

‘ALEXA, WHO ARE YOU?’ – ANALYSING ALEXA’S, CORTANA’S AND SIRI’S VOCAL PERSONALITY

Anabell Hacker

*Technische Universität Berlin, FG Kommunikationswissenschaft
anabell.hacker@tu-berlin.de*

Abstract: The paper answers the research question of whether listeners assign a personality to synthetic voices by using the example of Alexa by Amazon, Cortana by Microsoft and Siri by Apple. Moreover, the voices were assessed regarding their attractiveness. A perception experiment was conducted with 37 participants (28 female, nine male). Listeners were asked to rate the voice recordings using the 30-item short version of the NEO-FFI personality test and additionally on a seven-point attractiveness scale. The participants were able to assign personality to the synthetic voices. In the dimension of Extraversion, Siri showed the highest values, followed by Alexa. Alexa was deemed the most conscientious, with Siri and Cortana after that. Cortana had the highest Neuroticism values, followed by Siri and then Alexa. Siri was rated the most open to experience while Cortana received the lowest scores. Siri was also perceived as the most agreeable, Cortana had lower values in this dimension with Alexa in between. Siri and Alexa were rated significantly more attractive than Cortana. Lastly, an acoustic analysis of the recordings showed parallels to observations from research on the subject of voice and personality.

1 Introduction

Smart speech assistants are becoming more popular because of their ease of use and by conducting tasks automatically just by voice command [1] [2]. Due to the growing acceptance and use of these artificial intelligences, this paper answers the research question of whether users assign a personality to synthetic voices.

The question about the vocal personality of speech assistants has not yet been investigated. Previous research in the field of voice assistants has been limited to social science aspects, e. g. about making everyday life easier with virtual assistants [1] [3] or the consequences of artificial intelligence for society [4] [5] and data security [6]. In addition, a large area of research for virtual assistants lies in dialogue structure and its improvement [7] [8]. Research is also continuing on speech synthesis, in particular to improve and evaluate the output quality [9] [10], but more recently also in relation to emotions, i.e. how the systems recognise their users’ emotions and can react adequately [11] [12].

One publication that combines personality and voice assistants, however, examines to what extent the user’s personality influences the preference for one of the systems [13]. This study analysed Siri, Cortana and Google Now. It also examined the perceived attractiveness of the systems. Overall, Siri achieved the highest levels of attractiveness among the test subjects, followed by Google Now. Cortana landed in last place in the attractiveness rating.

2 Methodology

The voice assistants were recorded in Germany with their respective German voices. For the recordings two different types of texts were chosen. First, the voice assistants were asked for the weather forecast in Berlin by asking them ‘Wie wird morgen das Wetter?’. The systems answered the following:

Alexa: *‘Morgen ist das Wetter in Berlin teilweise sonnig mit einer Höchsttemperatur von 23 Grad und einer Tiefsttemperatur von 14 Grad.’*

Cortana: *‘Die Wettervorhersage für morgen meldet stellenweise klaren Himmel bei Temperaturen zwischen 17 und 23 Grad.’*

Siri: *‘Die Höchsttemperatur morgen wird 23 Grad Celsius und die Tiefsttemperatur 13 Grad Celsius sein.’*

These types of texts were chosen because a weather forecast is a common interaction with the systems by users. Secondly, the assistants needed to speak the same text in order to conduct comparable acoustic measurements. For these recordings four sentences from the Berlin Database of Emotional Speech [14] (short: Emo database) were chosen because these align well with the typical communication between a voice assistant and the user. The four sentences were: *‘Der Lappen liegt auf dem Eisschrank’*, *‘Das will sie am Mittwoch abgeben’*, *‘Heute Abend könnte ich es ihm sagen’* and *‘In sieben Stunden wird es so weit sein’*.

The recordings were conducted on July, 11th 2017 to maintain the same weather forecast and state of the voice assistants’ software. Alexa was recorded using an Amazon Echo Dot, second generation, Cortana on a Windows 10 PC with the Win 10 version 1703 and for Siri’s recordings an iPhone 6 with iOS 10.3 was used. The recordings were made with the software Audacity, version 2.3.1 [15].

During recordings, several issues occurred. Siri provides a built-in function to re-read text which has been highlighted by the user. The user can chose Siri’s reading speed which was set to 50 per cent to attain a medium speaking rate. Unfortunately, it was not possible to record Siri’s output via Bluetooth to avoid loss of quality due to digital to analogue conversion. Despite multiple tries using different software, Siri had to be recorded using an Aux cable. This resulted in a quiet background noise which is audible especially in between utterances. Alexa could not be recorded using Bluetooth either. It was possible to establish a Bluetooth connection to a PC and transfer the sound but as soon as the recording was started, Alexa stopped talking. Therefore, Alexa was also recorded with an Aux connection but without any noticeable loss in quality. Cortana’s recordings were conducted digitally using the internal audio of the PC.

Cortana and Alexa do not come with a user applied text-to-speech function. As a workaround an event was saved to the user’s calendar stating the Emo sentences as its heading. Thus, the systems read out the intended sentences when Alexa and Cortana were asked about the events. Later on, the date and time were cut out of the recordings using the software Praat [16].

3 Auditory Analysis

In order to find remarkable voice and speech features, the author conducted an auditory analysis of Alexa’s, Cortana’s and Siri’s recordings. As the way of speaking varies by the type of utterance – weather forecast and read-out text – the two recordings will be called *Alexa_W*, *Cortana_W* and *Siri_W* for the weather forecasts and *Alexa_Emo*, *Cortana_Emo* and *Siri_Emo* for the Emo database sentences.

Overall, the recordings *Alexa_W*, *Alexa_Emo* and *Siri_W* have the highest quality of naturalness. *Cortana_W* also sounds natural, but artifacts of the synthesis reduce the quality. The Emo recordings by Siri and Cortana have the lowest naturalness, with Siri slightly exceeding Cortana due to nicer accentuation. Cortana has the highest voice of the three assistance systems. Nonetheless, they are all on the lower end of average female voices. All systems articulate very precisely although Cortana speaks a little too precisely. In addition, the systems combine a relatively natural sentence declination, but little pitch variation within stressed syllables.

4 Perceptive Analysis of Alexa's, Cortana's and Siri's Vocal Personality

A hearing experiment was conducted in order to test on the following hypotheses:

Hypothesis 1: It is possible to assign personality to the virtual voice assistants Alexa, Cortana and Siri based on their voices and speech.

Hypothesis 2: It is possible to evaluate the attractiveness of the assistant systems based on their voices and speech.

Hypothesis 3: The personality assessment based on the Emo recordings and on the weather forecasts of the three systems is different.

Hypotheses 4: The reasons for the different personality perceptions of Alexa, Cortana and Siri are differences in (i) the mean pitch, (ii) stress, (iii) speaking rate, (iv) articulatory accuracy and (v) spectral properties of the voices.

4.1 Procedure

In order to analyse Alexa's, Cortana's and Siri's perceived personality the 30-item abridged version of the NEO-FFI in German according to Körner et al [17] was used. The listeners were asked to evaluate all six recordings using the NEO-FFI and on a seven item Likert scale (1: very unattractive, 7: very attractive) regarding the attractiveness of the voices.

37 listeners took part in the test with 28 being female and nine being male. All of them were German native speakers. The participants spread over various age groups: 15 were between 21 and 30 years old, six between 31 and 40 years old and one participant was in the group of 41 to 50 years-olds. The second most participants with a total of 13 listeners came from the group of 51 to 60 years-olds and two participants were older than 60.

Participants needed 22 minutes on average to complete the listening test. It was conducted online at SoSciSurvey.com. Participants were not told that they are hearing the voice assistants by Amazon, Windows and Google but only told that they are rating synthetic voices.

4.2 Results

As there weren't any studies on the Big Five personality of the three systems, two-sided hypotheses were used. In order to test on the hypotheses three paired t-tests (Alexa-Cortana, Alexa-Siri and Cortana-Siri) were conducted per Big Five dimension using the software R. To avoid Alpha error accumulation p was Bonferroni-corrected from .05 to .017 which results in a confidence interval of 98.3 per cent. For testing the homogeneity of the group of participants the intra class correlation (ICC) [18] was used. The listener group was checked for their ICC for each personality dimension of the NEO-FFI and reduced by as many listeners until the ICC comes as close as possible to 1 (100 per cent) without the ICC difference between the listeners falling below 0.001.

In the Extraversion rating of the Emo recordings 16 out of 37 listeners' ratings remained, for the weather forecast 22 out of 37 ratings remained. The assessment of Conscientiousness was calculated with nine (Emo) and ten judgments (weather). The data for Neuroticism contained 18 listener ratings for the Emo recordings and eleven ratings for the weather report. The dimension of Openness to Experience was rated by 20 (Emo) and 28 listeners (weather). In terms of Agreeableness, 23 of the 37 judgments for the Emo sentences were included in the t-tests. For the weather report, 29 of the 37 listeners' ratings remained. The attractiveness of the recordings in the Emo database was judged to be the sufficiently homogeneous by 30 listeners, and 32 listeners for the weather report.

It is important to note that the ICC does not measure the quality of the judgments, i.e. it does not show how appropriate the personalities attributed are to the actual personality of the speaker, but only measures the consistency of the hearers' judgments.

In the dimension Extraversion there was a significant difference between Alexa_Emo and Cortana_Emo ($p = .00003346$) as well as between Cortana_Emo and Siri_Emo ($p = .00003607$). Alexa_Emo and Siri_Emo did not differ significantly ($p = .175$). The recordings of the weather forecast only showed significant results for Alexa_W and Siri_W ($p = .000004109$) and for Siri_W and Cortana_W ($p = .00000003388$). The t-test did not return a significant difference between Alexa_W and Cortana_W ($p = .02801$). The means of the ratings are shown in Table 1, also in comparison to the norm sample by Körner et al [17].

In terms of Conscientiousness every t-test returned a significant result for the Emo recordings (Alexa_Emo > Cortana_Emo: $p = .00005046$; Alexa_Emo > Siri_Emo: $p = .005276$; Cortana_Emo < Siri_Emo: $p = .001503$). For the weather recordings, the difference was not significant for Alexa_Emo and Siri_Emo ($p = .025$).

Cortana was perceived the most neurotic with a significant difference to Alexa in both recordings (Emo: $p = .000002665$; weather: $p = .00004355$). In comparison to Siri, only the t-test with the weather forecast ratings returned a significant result ($p = .001267$). The Emo recordings of Cortana and Siri did not differ significantly ($p = .01738$). Alexa and Siri on the other hand did only show a significant difference in their Emo recordings ($p = .0001061$) but not in their weather forecasts ($p = .276$).

In the dimension Openness to Experience, there were significant results for both of Alexa's and Siri's recordings (Emo: $p = .0000318$; weather: $p = .006757$) and the comparison of Siri and Cortana (Emo: $p = .000008095$; weather: $p = .0000005284$). Although Siri was rated to have a higher level of Openness to Experience than Alexa in both recordings, only the weather forecasts received a significantly different rating ($p = .000007121$; Emo: $p = .709$).

The mean ratings for Alexa's and Cortana's Agreeableness differed significantly in both types of recordings (Emo: $p = .000000005314$; weather: $.000002638$) which was also true for the differences between Cortana and Siri (Emo: $p = .000000005486$; weather: $8.976 \cdot 10^{-11}$). Alexa_W and Siri_W also showed a significant difference ($p = .0006714$) but the t-test for Alexa_Emo and Siri_Emo did not return a significant result ($p = .239$).

In terms of attractiveness both of Alexa's and Cortana's recordings showed significant differences (Emo: $p = .00000004948$; weather: $p = .00000001295$). For Cortana and Siri there were also significant results when comparing both recordings: Emo: $p = .00000006889$; weather: $p = 5.504 \cdot 10^{-13}$. The differences in attractiveness between Alexa and Siri were not significant for the ratings of both recordings (Emo: $p = .935$; weather: $p = .121$).

Table 1: Means of the Personality, Attractiveness Ratings and the Norm Sample [17]

	Extraversion	Conscientiousness	Neuroticism	Openness to Experience	Agreeableness	Attractiveness
Alexa_Emo	2,01	3,02	0,96	2,19	2,30	3,97
Alexa_W	2,10	3,40	0,98	1,96	2,53	5,28
Cortana_Emo	1,23	1,70	2,04	1,20	1,12	1,90
Cortana_W	1,69	2,25	2,24	1,60	1,53	3,00
Siri_Emo	2,22	2,35	1,81	2,25	2,83	4,00
Siri_W	2,85	3,05	1,20	2,60	3,05	5,75
Norm sample	2,28	2,96	1,52	2,04	2,79	-

When comparing the systems in themselves between their Emo resp. weather recordings, only the attractiveness was always rated significantly different when comparing the Emo to the

weather values with the weather forecast always being more attractive than the Emo sentences. For the personality dimensions only five t-tests returned a significant result being the Extraversion of Siri_Emo and Siri_W ($p = .003492$), the Openness to Experience of Cortana's recordings ($p = .0001609$) and Siri's recordings (.008312) and Siri's Agreeableness ($p = .003573$).

5 Acoustical Analysis

Mainly the Emo recordings were used for the acoustical analysis because they provide the same wording for every system. The weather forecasts are used to point out measurements of hearable differences stated in the auditory analysis. Measurements were done using Praat [16].

As Praat shows a tendency to produce outliers when using its automatic pitch extraction, the median of the pitch will be regarded as the best fit to describe the voice assistants' mean pitches. When comparing Alexa's, Cortana's and Siri's pitches (s. Table 2) to the mean woman's pitch in Germany, which is located at 220 Hz [19], the perceived lower pitch can be confirmed.

Table 2: Pitch measurements in Hertz

	Arithmetic mean	Median	Standard deviation
Alexa_Emo	182	176	36
Cortana_Emo	193	190	48
Siri_Emo	170	172	17

The standard deviation of the pitch indicates how vivid speech is. Lower numbers show more monotonous speech whereas higher values indicate more vivid speech. Siri's standard deviation is comparable to the one of a news reporter (about 20 Hz), whereas Alexa's and Cortana's values are more equal to those found in TV hosts (about 40 Hz). The standard deviation was lower in Alexa's and Cortana's weather recordings (Alexa_W: 34 Hz, Cortana_W: 36 Hz) and higher in Siri_W (20 Hz).

In order to analyse Alexa's, Cortana's and Siri's way of sentence and word stress the pitch and duration of stressed and unstressed syllables were measured. Alexa uses mainly a raise of pitch to stress syllables, the duration of stressed and unstressed units is very similar. Cortana on the other hand, enables higher pitch as well as prolonged duration of stressed syllables. However, Siri seems to use longer durations to stress rather than pitch because her pitch is very similar in every level of stress. To describe the way of intonation more, the Praat script Prosogram [20] was used. The Emo recordings of the systems mainly show constant courses of intonation or even falling movements which backs the impression of monotonous or even bored speech [21]. The weather recordings however really show more movements in intonation. Especially in Alexa's weather forecast, more movement in stressed units can be observed. However, Siri's intonation is still the most monotonous with the most constant courses of intonation.

The speaking rate was measured in syllables per second. Cortana speaks the most slowly with 3.9 syllables/sec in the Emo recording and 4.5 syllables/sec during the weather forecast. Alexa follows shortly after Cortana with 4.3 syllables/sec (Alexa_Emo) and 4.8 syllables/sec (Alexa_W). Siri speaks at the highest rate with 4.7 for Siri_Emo and 4.9 for Siri_W. All systems spoke faster in their weather forecasts than during the Emo recordings. Still, all speaking rates are near to the average of 4.4 syllables/sec for read out text [22] or slightly over the rates for improvised speech (3.16 syllables/sec [22] or 4.17 syllables/sec [23]).

Alexa's, Cortana's and Siri's perceived high articulatory precision turned out to be true in parts. Most vowels show normal formant values as observed in average female German speakers [24]. Cortana reaches a vowel target overshoot in her [a], producing formant frequencies as high as 1015 Hz for the first formant (F_1) and 1611 for the second formant (F_2) (standard: F_1 : 836 Hz, F_2 : 1586 Hz [24]). Alexa and Siri on the other hand overshoot the vowel target for [i]

(Alexa: F1: 348 Hz, F2: 2652 Hz; Siri: F1: 327 Hz, F2: 2711 Hz; standard: F1: 302 Hz, F2: 2533 Hz). Moreover, their formant frequencies for [e] are very close to a standard [i] (Alexa: F1: 299 Hz, F2: 2651 Hz; Siri: F1: 343 Hz, F2: 2669 Hz; standard: F1: 434 Hz, F2: 2461 Hz).

When comparing the spectral slices of the /a/ in the word ‘Abend’ of the Emo sentences, Alexa’s strong and stable voice is proven by a lot of power throughout the whole spectrum. Siri’s spectral slice looks the most natural for the weather recording, in the Emo sentences however, the lower frequencies outweigh the higher ones which usually indicates sad or resigned emotion [25]. Due to the look of the oscillogram of the /a/, Alexa’s voice is darker as Cortana’s and Siri’s because the oscillation is more similar to a sine curve rather than impulse-like as Cortana’s and Siri’s oscillograms are.

6 Conclusion

Hypothesis 1 and 2 (s. chapter 4) could be verified by the evaluation of the listening experiment. None of the participants showed any difficulties assigning a personality to the synthetic voices. Moreover, significant differences were found between the personalities of the three voices which indicates that people notice individual traits. This is also applicable for attractiveness.

Hypothesis 3 that the ratings of the Emo and weather recordings differed from each other cannot be accepted in its general form. Although the quality of the synthesis were perceived differently for Cortana and Siri, these could not always be verified statistically. Only for the rating of attractiveness, the weather forecast was rated significantly more attractive than the Emo recordings throughout all systems.

As three recordings are too small of a dataset to find statistic correlations, the parts of hypotheses 4 could neither be verified nor falsified. In the following, the results and measurements are compared to previous findings of research [21]. It was replicated that higher mean pitch could be found in persons with higher levels of Neuroticism, Cortana in this case. That lower pitches in women’s voices are perceived as pleasant, trustworthy and competent [19] could be linked to Siri and Alexa being rated as more attractive than Cortana. Alexa, whose spectrum showed the most power in higher frequencies, was rated to have a low level of Neuroticism which was also a result of previous research [21]. Cortana, who was rated most neurotic, had the least power in higher frequencies in her spectral slice. Findings about acoustical correlates for Extraversion could not be replicated with the data on hand. Nevertheless, the order of some acoustic measurements were the same or reversed as the order found in most personality ratings (Siri > Alexa > Cortana). This sequence is applicable for the speaking rate, whereas the standard deviation of the pitch and the mean pitch show the reversed order because Cortana showed the highest values, followed by Siri and then Alexa. These findings could be used to generate more hypotheses for further research.

7 Discussion

In order to find acoustic correlates to the personality of the voice assistants, more data should be collected to conduct statistical tests of correlation. A general problem about working with the voices of the three systems is that the companies constantly operate on them. This means that the recordings can become obsolete in a very short matter of time. So research on them can only be a snapshot of their development.

Another weakness of the present data is the imbalance of participants in the listening experiment in regard to their sex. Also, there should be more data collected about the participants such as experience with voice assistants and their attitude towards them. But as the participants were not primed as to which voice assistants they were listening to, they might not even have recognised the synthetic voices presented. Still, regular users of Alexa, Cortana and Siri might

have been able to identify them. So it would be useful to ask participants whether they recognised the voice and use the assistant regularly.

All in all, it can be said that listeners do assign a personality to synthetic voices. This finding can be relevant in the development of voice assistants, especially for companies. It is probable that the personality a person assigns to the system, is transferred to the company providing the system [26] [27]. And as people tend to adapt to voice systems [28] and because voice-only communication elicits higher rates of empathic accuracy than vision-only or multidimensional communication [29], users might even be able form an emotional bond to these personalities if programmed and chosen well.

8 References

- [1] BOTH, G. , „Better Living Through Siri? Arbeitersparnis, Geschlecht und Virtuelle Assistent_innen,“ *Zentrum für transdisziplinäre Geschlechterstudien Bulletin Texte* 40, pp. 123-138, 2012.
- [2] HOLST, A., „statista.com“, 4 January 2021. [Online]. Available: <https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/>. [accessed 20 January 2021].
- [3] WECHSUNG, I., EHRENBRINK, P., SCHLEICHER, R. and MÖLLER, S., „Investigating the social facilitation effect in human-robot-interaction,“ in *Natural Interaction with Robots, Knowbots and Smartphones Putting Spoken Dialog Systems into Practice*, New York, Springer, 2014, pp. 167-177.
- [4] GEHL, R. W. and BAKARDJIEVA, M., *Socialbots and Their Friends: Digital Media and the Automation of Sociality*, New York: Routledge, 2017.
- [5] GUZMAN, A. L., „Making AI Safe for Humans: A Conversation with Siri,“ in *Socialbots and Their Friends: Digital Media and the Automation of Sociality*, New York, Routledge, 2017, pp. 69-85.
- [6] FERDINAND, J.-P., JETZKE T., „Voice Computing - allgegenwärtige Spracherkennung,“ in *Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag. Themenkurzprofil Nr. 15*, 2017.
- [7] SARIKAYA, R., CROOK, P. A., MARIN, A., JEONG, M., ROBICHAUD, J. P., CELIKYILMAZ, A. and KIM, Y. B., „An Overview of End-to-End Language Understanding and Dialog Management for Personal Digital Assistants,“ in *2016 IEEE Spoken Language Technology Workshop (SLT)*, San Diego (USA), 2016.
- [8] HILLMANN, S., WEISS, B., MICHAEL, T., MÖLLER, S., „Simulation von Nutzerverhalten bei der Interaktion mit Sprachdialogsystemen mittels Künstlicher Neuronaler Netzwerke,“ in *29. Konferenz Elektronische Sprachsignalverarbeitung (ESSV)*, Ulm, Studententexte zur Sprachkommunikation, Dresden, TUDpress, 2018, pp. 151-158.
- [9] HINTERLEITNER, F., *Quality of Synthetic Speech. Perceptual Dimensions, Influencing Factors and Instrumental Assessment.*, Dissertation at the TU Berlin, 2016.
- [10] NORRENBROCK, C. R., HINTERLEITNER, F., HEUTE, U. and MÖLLER, S., „Instrumental Assessment of Prosodic Quality for Text-to-Speech Signals,“ *IEEE Signal Processing Letters*, Vol. 19, No. 5, pp. 255-258, Mai 2012.
- [11] BURKHARDT, F. and WEISS, B., „Complex Emotions - The Simultaneous Simulation of Emotion-Related States In Synthesized Speech,“ in *28. Konferenz Elektronische Sprachsignalverarbeitung*, Dresden, TUDpress, 2017, pp. 67-74.
- [12] WU, B., JIA, J., HE, T., DU, J., YI, X. and NING, Y., „Interferring User's Emotions for Human-Mobile Voice Dialogue Applications,“ in *2016 IEEE International Conference on Multimedia and Expo (ICME)*, Seattle (USA), 2016.

- [13] EHRENBRINK, P., OSMAN, S. and MÖLLER, S., „Google Now is for the Extraverted, Cortana for the Introverted: Investigating the Influence of Personality on IPA Preference,“ in *Proceedings of the 29th Australian Conference on Human-Computer Interaction, Brisbane, December 2017 (OZCHI'17)*, Texas Australia, 2017.
- [14] BURKHARDT F., PAESCHKE A., ROLFES M., SENDLMEIER W. F. and WEISS B., „A Database of German Emotional Speech,“ in *Interspeech'2005 - Eurospeech, 9th European Conference on Speech Communication and Technology*, Lissabon, 2005.
- [15] AUDACITY TEAM, „Audacity, Version 2.3.1,“ 2017. [Online]. Available: <https://www.audacityteam.org/>. [accessed 20 January 2021].
- [16] BOERSMA, P. and WEENINK, D., „Praat, Version 6.0.36,“ 2017. [Online]. Available: <http://www.fon.hum.uva.nl/praat/>. [accessed 20 January 2021].
- [17] KÖRNER, A., GEYER, M., ROTH, M., DRAPEAU, M. et al., „Persönlichkeitsdiagnostik mit dem NEO-Fünf-Faktoren-Inventar: Die 30-Item-Kurzversion,“ *Psychotherapie, Psychosomatik, Medizinische Psychologie*, Bd. 58, pp. 238-245, August 2008.
- [18] BORTZ, J. and DÖRING, N., *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*, 4., überarbeitete Auflage, Heidelberg: Springer Medizin Verlag, 2006.
- [19] SENDLMEIER, W. F., „Die psychologische Wirkung von Stimme und Sprechweise,“ in *Resonanz-Räume. Die Stimme und die Medien.*, Berlin, Bertz + Fischer, 2012, pp. 99-116.
- [20] MERTENS, P., „Prosogram, Version 2.14d,“ 24 January 2018. [Online]. Available: <https://sites.google.com/site/prosogram/download>. [accessed 09 February 2018].
- [21] SENDLMEIER W. F., *Sprechwirkungsforschung. Grundlagen und Anwendungen mündlicher Kommunikation*, 3. korrigierte Auflage, Mündliche Kommunikation, Band 10, W. F. Sendlmeier, ed., Berlin: Logos, 2019.
- [22] GEIßNER, G., *Mündlich: Schriftlich. Sprechwissenschaftliche Analysen "freigesprochener" und "vorgelesener" Berichte*, Frankfurt am Main: Scriptor, 1988.
- [23] TRÄGLER, D., *Der Einfluss der Synchronstimme auf die Charakterdarstellung*, Mündliche Kommunikation, Bd. 9, W. F. Sendlmeier, ed., Berlin: Logos Verlag, 2014.
- [24] SENDLMEIER, W. F. and SEEBODE J., „Formantkarten des deutschen Vokalsystems,“ 2007. [Online]. Available: https://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf. [accessed 23 April 2018].
- [25] PAESCHKE, A. and SENDLMEIER, W. F., „Die Reden von Rudolf Scharping und Oskar Lafontaine auf dem Parteitag der SPD im November 1995 in Mannheim - Ein sprechwissenschaftlicher und phonetischer Vergleich von Vortragsstilen,“ *Zeitschrift für Angewandte Linguistik*, Nr. 27, pp. 5-39, 1997.
- [26] ASIA RADIO TODAY, "The personality of voice assistants" on „[asiaradiotoday.com](http://www.asiaradiotoday.com),“ 21 June 2020. [Online]. Available: <http://www.asiaradiotoday.com/news/personality-voice-assistants>. [accessed 18 January 2021].
- [27] OHAYON, P., "Trademark your brand's speech pattern" on „phoebeohayon.com,“ 2020. [Online]. Available: <https://phoebeohayon.com/2019/09/07/trademark-your-brands-speech-patterns/>. [accessed on 18 January 2021].
- [28] COHN, M. D., SEGEDIN, B. F. and ZELLOU, G., „Imitating Siri: Socially-Mediated Vocal Alignment to Device and Human Voices,“ in *International Congress of Phonetic Sciences (ICPhS)*, Melbourne, Australia, 2019.
- [29] KRAUS, M. W., „Voice-only communication enhances empathic accuracy,“ *American Psychologist*, Nr. 72(7), pp. 644-654, October 2017.