Netnography of Social Media Addresses on COVID-19

Vered Aharonson[1,2], Christos Karpasitis [2], Taliya Wenstein[1] and Gershon Koral[1]
[1]School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa
[2]School of Sciences, University of Central Lancashire, Pyla, Cyprus
Vered.aharonson@wits.ac.za
ckarpasitis1@uclan.ac.uk
tali.jae@gmail.com
gershonkoral@gmail.com

Abstract: Healthcare professionals' harness social media to encourage responsible behaviour during the COVID-19 pandemic. As internet users often struggle assessing the veracity of the information in these addresses, acoustic characteristics of the presenters' speech may play a significant role in their persuasiveness impact. Using a netnographic approach, we studied YouTubers' reactions to explore the persuasiveness attributes of COVID-19 related speeches included in YouTube videos within a South Africa context. The persuasiveness index was computed from the view count, likes and dislikes of 314 speech segments from YouTube interviews related to COVID-19. Standard acoustic features – Mel frequency cepstral coefficients - of the interviewees' voice were extracted through speech processing. Recurrent neural networks were optimized and evaluated the strength of these acoustic features to classify and predict the persuasiveness index. The cepstral feature set yielded a balanced accuracy of 86.8% and F1 score of 85.0%. These preliminary results exhibit the potential of the vocal cepstrum as predictor of persuasiveness in healthcare addresses on responsible behaviour during the COVID-19 pandemic. The results imply that quantitative acoustic analysis of a presenter's voice, independent from text, can explain the impact of social media addresses.

Keywords: COVID-19, Public address, Persuasiveness, YouTube interviews, Speech Analytics, Deep Machine Learning

1. Introduction

Public addresses have the potential to promote behaviour which may limit the spread of COVID-19, such as wearing masks, keeping social distance and washing hands (Kollamparambil and Oyenubi, 2021). Social media are important channels to exploit this potential and to encourage these behavioural changes on a large scale, as digital space has become a focal point where communities access and interact with information and communication (Park et al., 2014). Furthermore, social media offers practical methods to study these healthcare-related addresses and to assess their effectiveness on large populations (Fenton and Procter, 2019). Social network analysis (SNA) is an important component of netnography (internet ethnography), where communities can be probed (Kozinets, 2015). Ampofo (2011) describes netnography as studying online communities and utilising their publicly available information to identify their needs and desires. Moreover, this research method can provide online data and deeper insights into consumers' opinions, motives, and concerns (Orgad, 2009; Kozinets, 2010). while also enabling comprehensive and ethical access to online community members and an insight into relevant consumer opinions and attitudes (Kozinets, 2017). In comparison to participant-driven research, this method can provide access to communities as large as nations (Fenton and Procter, 2019).

As a research method, netnography is primarily employed in research related to marketing (Kozinets, 2015). Karpasitis et al. (2019) also used netnography to study the engagement of Social Media users with online video advertisements posted on Youtube. Beyond marketing and advertising, however, some attention has also been given to the use of social media for social causes and human betterment (Kanter and Fine, 2010). In the past decades more national and international public addresses, for social causes have been uploaded on social media. Promoting responsible behaviour in the time of the pandemic can be regarded as one of these causes. In the past two years of the COVID-19 pandemic, a prolific number of addresses strive to alert and persuade people to adopt responsible and safe behavioural measures.Yet, both unfamiliarity and controversies exist in the scientific and medical communication available online dealing with COVID-19 prevention measures. This pertains to even the elementary recommendations for safe behaviour, such as wearing masks, washing hands and social distancing.

As in marketing of products/services, the persuasiveness in social media addresses on social causes, is greatly governed by the attributes of the message delivery. The studied community rarely possesses the means to

ascertain the veracity of the message content. Thus, a community's reactions on "likes" and "dislikes" often reflects opinions on the manner the message was conveyed or the person who conveyed it. The providers of theses addresses and their delivery can therefore benefit from a quantitative evaluation of the non-verbal attributes of their message's delivery.

In this study we took a test-case of a South African Youtube community, aiming to quantify the non-verbal attributes of persuasive addresses on COVID-19 safe-behavioural measures. Preliminarily we focused on the acoustic attributes of the delivery, namely the vocal features of the delivering person. The acoustic attributes are easier to extract compared to visual attributes which vary significantly between addresses. Some addresses portray images and video clips whereas others focused the camera on the speaker, some addresses display only a speaker's face, while other portray a whole body, and more. Moreover, voice-only analytics can be transferred unto multiple delivery channels, where no visual attributes exist, such as audio podcasts.

Most netnographic studies employ qualitative analysis, and where quantitative analysis is pursued, mainly manual coding schemes are utilised (Fenton and Procter, 2019). The digital space where the netnographic research takes place, however, opens a vista of opportunities to employ digital processing tools, such as artificial intelligence, to expedite and improve the analysis.

Our study harnessed speech processing and machine learning tools to quantitatively analyse and evaluate voice properties of public addresses on COVID-19. Our aim was to quantify a persuasiveness measure for public addresses on COVID-19 and to use it for the assessment of the persuasiveness attributes in the addresses. Although persuasiveness attributes can be multi-modal, i.e. visual, verbal and acoustic properties, (Nojavanasghari et al., 2016) our preliminary study focuses on the acoustic properties. We harnessed signal processing to quantify these properties and machine learning to evaluate their contribution to the persuasiveness measure. The analysed interviews are derived from the South African based news outlet eNCA's YouTube channel. Thus, the persuasiveness index is assumed to capture the reactions of a majority South African based demographic. South Africa has a social media user base of 25 million people (Kemp, 2021) with 73.5% of users aged between 18-44 with a roughly even split between male and female users (Kemp, 2021).

2. Methods

As mentioned earlier, the netnographic research approach of selecting publicly available online videos and analysing their content characteristics and their impact to viewer's repsonses was also used successfully by other authors in the past. This makes this research method the most appropriate way to undertake this study. The following sections explain how processes such as speech data collection, persuasiveness measure, speech processing and machine learning were used alongside this research approach.

2.1 Speech Data Collection

To reduce variability between types of addresses and social media platform, a YouTube channel of a popular news channel in South Africa, eNCA, was selected. This choice also confined possible variability due to the language and accent differences in South African English. This simplifies the preliminary analysis sought in our study.

The data included 38 interview videos of reputable South African interviewees on non-pharmaceutical COVID-19 behaviours in line with those provided by the World Health Organisation (WHO) (Organization, 2020). The interviews were sourced from the eNCA YouTube channel, upon permission of YouTube's data extraction policy (Acker and Kreisberg, 2020). A ratio of 9:10 for women to men interviewees ensured equal gender representation. A screenshot of an interview is portrayed in figure 1.

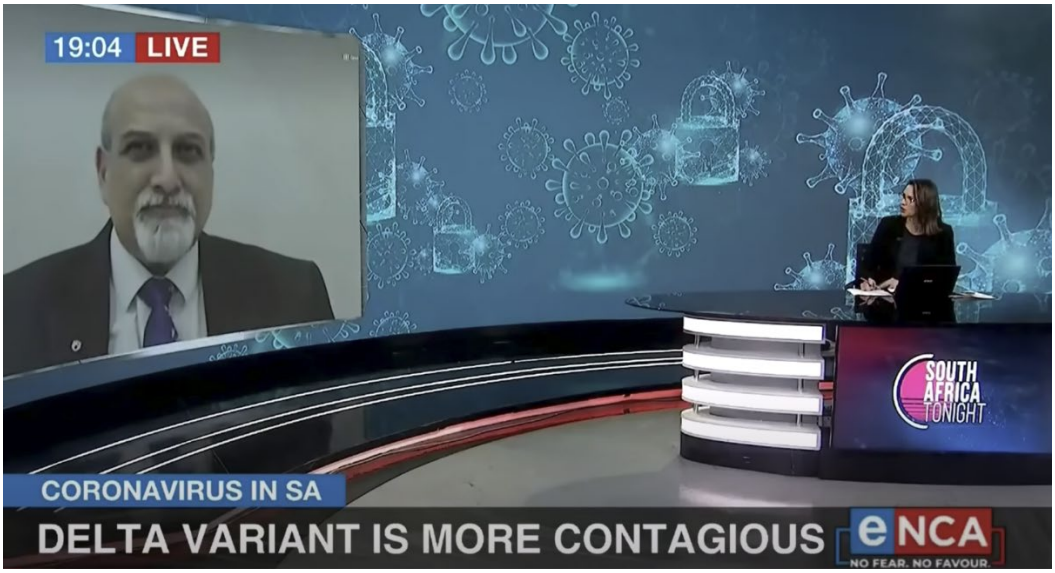

Figure 1: An example screenshot of an interview video from the eNCA YouTube channel

2.2 Netnographic Persuasiveness Measure

$$L = a + \frac{(b-a)(X - X_{min})}{X_{max} - X_{min}} \quad (1)$$

where $X = \frac{views \cdot likes}{dislikes}$ and a and b are the normalisation limits set to -1 and 1, respectively.
The labels(L) were then binarized to provide two classification categories: persuasive and non-persuasive, using a threshold of m±∑, where m is the median of the dataset, which reduces class imbalance, and ∑, set to 0.05 to compensate for the right-skewed data distribution.

2.3 Speech processing

Speaker diarisation was performed to separate the interviewees speech segments from the interviewers'. Thereafter, the segments were clipped to 30s sequences to ensure uniform lengths. Segments shorter than 30s were discarded. This segmentation was expedient for the signal processing algorithms employed later. The segmentation created more data sequences and therefore provided augmentation to increase the data size. The audio sequences were pre-processed and their features were extracted using standard emotive speech processing for machine learning (Eyben, 2016). A low-pass filter, with a 17 kHz cutoff frequency was applied to limit the analysis to a frequency range audible to the adult human ear (Monson and Caravello, 2019). Noise removal was performed by auto-spectral subtraction with a voice preset. A hamming window was used to split the clips into 30ms frames with 50% overlap.

2.4 Feature Extraction

The features extracted were Mel Frequency Cepstral Coefficients (MFCC), which are widely used in numerous speech processing tasks. Specifically, MFCC were successful features for speaker traits recognition and in emotion recognition from speech. This feature set is based on models of the human hearing system, is robust to noise and is relatively consistent across genders (Venkataramanan and Rajamohan, 2019). The first 13 MFCC together with their first and second derivatives were extracted. Each audio clip, thus, had a feature set of 39 features with each feature standardized to have a zero mean and a standard deviation of 1 to assist with the learning rates of the machine learning models.

2.5 Machine learning

The main challenge for deep machine learning was the smallness of our dataset. Our machine learning methods therefore put emphasis on over-fitting reduction. These overfitting reduction strategies include L2 regularisation, drop out and early stoppage with a validation patience of 5. The dataset was split into a training, validation and test set, in ratios of 12:5:3, respectively. A recurrent neural network (RNN) model, commonly used for speech analytics, was designed to predict the persuasiveness measures based on the audio feature set. The model architecture was determined by a Bayesian optimization strategy. The models' validation was performed every 50 iterations using a holdout method. The investigated model employs a cross entropy loss function and ADAM optimizer, with a momentum of 0.9 (Nojavanasghari et al., 2016, Ravanelli et al., 2018). This

optimal model was re-run 3 times with different seed values, for 30 iterations of the Bayesian optimization. Thereafter, the best performing seed value was evaluated over 5 randomly assigned runs to generate a statistical spread of model performance to account for the different data divisions in the small dataset.

Balanced accuracy - to mitigate class imbalance bias, and F1 scores – to identify false positive and negative predictions, were used to evaluate the model performance (Thanaraj et al., 2021). The validation loss was added to the performance metrics to evaluate the over-fitting effect. The metrics of the best performing seeded model, as well as the mean performance metrics for the 5 randomly split video models were considered in the analysis.

3. Results

The data collection yielded a dataset of audio clips as summarized in Table 1.

Table 1: Dataset content

Number of clips in the dataset			
	Male speakers	Female speakers	Total
Persuasive	104	33	137
Non-persuasive	71	106	177
Total	175	139	314

The set of persuasiveness labels, L, computed on the set of clips had a median m of 0.775. This value was used to binarize the labels.

The architecture of the RNN which achieved the highest validation accuracy and minimal validation loss among the 5-fold models is depicted in figure 2.

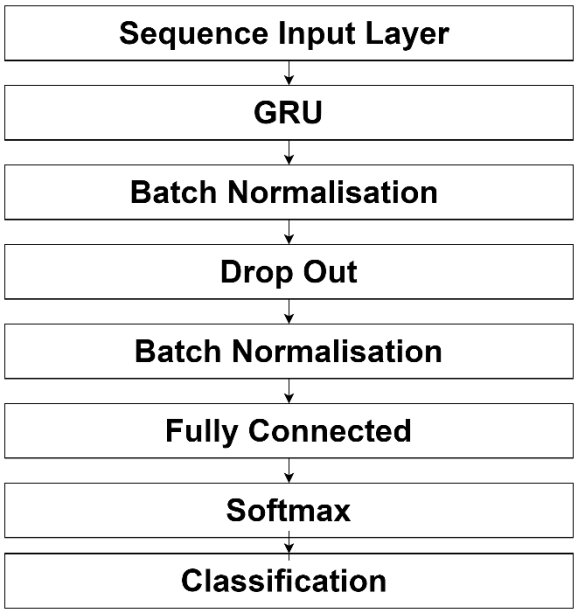

Figure 2: The model architecture of the best performing RNN

The hyperparameters of the best performing RNN model are listed in Table 2.

Table 2: Hyperparameters of the best performing RNN

Initial Learning Rate	Hidden Units	Drop Out	Epochs	Batch	L2 Regularization
0.0908	5	0.5518	271	34	0.0114

Table 3 illustrates the performance metrics, for the best performing model and for the ensemble of the 5 models generated in the 5-fold validation procedure.

Table 3: Persuasiveness prediction performance metrics

	balanced accuracy (%)	F1 score (%)	validation loss (%)
Best Model (Same seed)	86.8	85.0	27.4
5 best models [mean±SD] random split	79.1±4.3	76.0±5.6	52.8±13.0

4. Discussion and conclusions

A long-standing research debate strives to assess the balance between cons and pros of social media. Quantifying the persuasiveness characteristics of messages delivered on social media and their effect on populations can assist in the analysis of this debate.

Our study investigated a South African based YouTube community through their reactions (in the form of "like" and "dislike") to interview content to develop a persuasiveness index for public addresses on COVID-19. This preliminary study is the first investigating COVID-19 addresses persuasiveness through nethnography. Previous netnographic studies were mostly concerned with persuasiveness in marketing applications. Moreover, many other studies employed an explicit ground-truth persuasiveness score, by recruiting participants, focus groups or "Mechanical Turk" methods. Yet, our study quantitatively analysed the speakers' attributes that contribute to this persuasiveness index, using speech processing and machine learning. The performance outperformed previous works which used comparable computational analysis, i.e. Park, S., et al. (2014) who achieved a mean accuracy of 66% at predicting persuasion using acoustic features (Park et al., 2014). Our best model yielded a balanced accuracy of 86.8%, and an F1 score of 85.0%. The results convey that over-fitting was successfully reduced, as reflected in a validation loss of 27.4% in the single test case analysis. The results imply that the cepstral set contains robust features for persuasiveness prediction and demonstrates the promise of utilising RNN in persuasiveness analysis. The optimization strategy yielded a best performing model architecture consisting of GRU, L2 regularisation and dropout layers. These components were previously indicated to outperform other RNN models in small audio datasets (Shewalkar et al., 2019). The batch normalization in this model was similarly indicated previously as improving system performance through gradient stabilization (Ravanelli et al., 2018).

Two assumptions in our study should be noted. The first is that the like-to-dislike index defined in the study is a reliable measure for persuasiveness. A scale of "like-to-dislike" was used in traditional studies that used recruited subjects who rated persuasive speech (Edwards and Von Hippel, 1995) and in more recent studies where this scale was used for social software persuasion (Broekens and Brinkman, 2009). The second assumption is that the single modal of acoustic properties of speech is sufficient to capture persuasiveness in the addresses on COVID-19, although the YouTubers' who watched these addresses were exposed to the textual content of speech as well as to visual attributes of the speaker. The textual content is a fundamental component in speech. Persuasion and emotion recognition studies noted, however, the major importance and sometimes dominance of the acoustic over the textual cues (Chuang and Wu, 2004). Moreover, in many cases social media users cannot ascertain the veracity of the textual content in these addresses. Specifically, many YouTubers are not familiar with and could not understand the scientific and medical messages pertaining to COVID-19. This causes a lack of trust in the content of such addresses (van Dijck and Alinejad, 2020, Malecki et al., 2021). Improving the voice attributes of the speakers has a potential to enhance the trust of social media users in these addresses.

Our study focused on a South African news network YouTube channel and monitored South African YouTubers' and speakers' reactions to explore the persuasiveness attributes of COVID-19 related speeches included in YouTube videos within a South Africa context. The development of a persuasiveness index for public addresses on COVID-19 was this study's main objective and it is considered accomplished. To this end, our study's contribution to the field is twofold. Firstly, the optimization strategy and the methods used for data collection and analysis support previous literature which also demonstrate the benefits of 5tilizing RNN in persuasiveness analysis. Moreover, our preliminary results exhibit the potential of the vocal cepstrum as predictor of persuasiveness in healthcare addresses on responsible behaviour during the COVID-19 pandemic while also implying that quantitative acoustic analysis of a presenter's voice, independent from text, can explain the impact of social media addresses. This could act as a motivation for the development of larger future studies in this field,, with international data, that would be able to examine the generalization of the methods and results proposed in this study to other ethnicities and populations. Further, the study focused on addresses on COVID-

19 safe behaviour, a natural research content in the time of the pandemic. The study could be extended, however, to examine other social causes that affect our lives and other social media addresses to these causes.

To conclude, although preliminary, our study offers additional insight into the mechanisms behind online persuasion, within the important context of behaviour change during the COVID-19 pandemic.

References

Acker, A. & Kreisberg, A. 2020. Social media data archives in an API-driven world. *Archival Science,* 20**,** 105-123.

Ampofo, L. 2011). The Social life of real-time Social Media monitoring. Journal of Audience and Reception Studies, Volume 8, Issue 1 (May 2011), pp 21-47.

Broekens, J. & Brinkman, W.-P. Affectbutton: Towards a standard for dynamic affective user feedback. 2009 3rd international conference on affective computing and intelligent interaction and workshops, 2009. IEEE, 1-8.

Chuang, Z.-J. & Wu, C.-H. Multi-modal emotion recognition from speech and text. International Journal of Computational Linguistics & Chinese Language Processing, Volume 9, Number 2, August 2004: Special Issue on New Trends of Speech and Language Processing, 2004. 45-62.

Edwards, K. & Von Hippel, W. 1995. Hearts and minds: The priority of affective versus cognitive factors in person perception. *Personality and Social Psychology Bulletin,* 21**,** 996-1011.

Eyben, F. 2016. Acoustic features and modelling. *Real-time Speech and Music Classification by Large Audio Feature Space Extraction.* Springer.

Fenton, A. & Procter, C. 2019. Studying social media communities: blending methods with netnography. *SAGE Research Methods Cases*.

Kanter, B. & Fine, A. 2010. The networked nonprofit: Connecting with social media to drive change, John Wiley & Sons.

Karpasitis, C., Polycarpou, I., Kaniadakis, A., Kvasova, O. 2019. I See, I Hear, I Speak – How Audiovisuals Affect Brand Video Virality. International Conference on Research in Advertising 2019, Krems, June 2019.

Kemp, S., 2021. *Digital in South Africa: All the Statistics You Need in 2021 — DataReportal – Global Digital Insights*. [online] DataReportal – Global Digital Insights. Available at: https://datareportal.com/reports/digital-2021-south-africa [Accessed 22 January 2022].

Kollamparambil, U. & Oyenubi, A. 2021. Behavioural response to the Covid-19 pandemic in South Africa. *Plos one,* 16**,** e0250269.

Kozinets, R. 2017. Netnography: Radical participative understanding for a networked communications society. The SAGE handbook of qualitative research in psychology, 374.

Kozinets, R. V. 2010. Netnography: Doing Ethnographic Research Online, SAGE.

Kozinets, R.V. 2015. *Netnography: redefined*, Sage

Malecki, K.M., Keating, J.A. & Safdar, N. 2021. Crisis communication and public perception of COVID-19 risk in the era of social media. Clinical Infectious Diseases, 72, 697-702.

Monson, B.B. & Caravello, J. 2019. The maximum audible low-pass cutoff frequency for speech. The Journal of the Acoustical Society of America, 146, EL496-EL501.

Nojavanasghari, B., Gopinath, D., Koushik, J., Baltrušaitis, T. & Morency, L.-P. Deep multimodal fusion for persuasiveness prediction. Proceedings of the 18th ACM International Conference on Multimodal Interaction, 2016. 284-288.

Orgad, S. 2009. How can researchers make sense of the issues involved in collecting and interpreting online and offline data?. In A. N. Markham and N. K. Baym (eds), Internet Inquiry: Conversations About Method. Thousand Oaks: Sage (pp. 33-53).

Organization, W.H. 2020. Coronavirus disease 2019 (COVID-19): situation report, 73.

Park, S., Shim, H.S., Chatterjee, M., Sagae, K. & Morency, L.-P. Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach. Proceedings of the 16th International Conference on Multimodal Interaction, 2014. 50-57.

Ravanelli, M., Serdyuk, D. & Bengio, Y. 2018. Twin regularization for online speech recognition. *arXiv preprint arXiv:1804.05374*.

Shewalkar, A., Nyavanandi, D. & Ludwig, S.A. 2019. Performance evaluation of deep neural networks applied to speech recognition: RNN, LSTM and GRU. *Journal of Artificial Intelligence and Soft Computing Research,* 9**,** 235-245.

Thanaraj, K.P., Noel, J.R.A. & Vijayarajan, R. 2021. Emotion classification from speech signal based on empirical mode decomposition and non-linear features. *Complex & Intelligent Systems,* 7**,** 1919-1934.

Van Dijck, J. & Alinejad, D. 2020. Social media and trust in scientific expertise: Debating the Covid-19 pandemic in the Netherlands. Social Media+ Society, 6, *2056305120981057.*

Venkataramanan, K. and Rajamohan, H., 2019. Emotion Recognition from Speech. ArXiv, abs/1912.10458.

Weber, K., Tinnes, L., Huber, T., Heimerl, A., Reinecker, M.-L., Pohlen, E. & André, E. Towards demystifying subliminal persuasiveness: using XAI-techniques to highlight persuasive markers of public speeches. International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems, 2020. Springer, 113-128.