Application for ethical approval at DTU Compute's Institutional Review Board

Project title:

DanskEmoTale (DanishEmoSpeech): a pilot study

Principal investigator:

Line Katrine Harder Clemmensen, Ph.d., associate professor, DTU Compute

Other investigators:

Sneha Das, Postdoc, DTU Compute

Date: 14.12.2021

Contact person: Sneha Das (sned@dtu.dk)

1 Project background, objectives and hypotheses

1.1 Background

Speech emotion recognition (SER) refers to the technique of studying and interpreting the emotional state of an individual from acoustic cues. With a rich body of literature since the late 1990s, the topic continues to garner interest, more so with recent applications in speech technology, multimodal affective computing and in healthcare [1, 2]. SER combined with affect recognition, which uses other modalities like vision and physiological signals are deployed in a wide range of applications. For instance, in the detection and intervention of disorders in healthcare, monitoring the attentiveness of students in schools, risk assessment within the criminal justice system, and for commercial applications, like detecting customer satisfaction in call-centers and by employment agencies to find suitable candidates [3, 4]. Research in SER requires access to speech corpora annotated with emotions. While multiple corpora for commonly spoken languages exist, there is a lack of functional Danish speech emotion data sets [5]. Unavailability of Danish affect data sets not only impedes the development of the technology, but also impacts the validation of existing methods on Danish speakers. This project is initiated with the broad objective to narrow the gap in speech affective computing within Danish speakers by developing a Danish speech affect data set.

1.2 Objective

The broad goal of this project is to enable studies on speech emotion recognition for Danish speakers. Therefore, the project will pilot the creation of a Danish speech affect data set [6]. The developed corpus will comprise of a Danish speech recording and the associated affect annotations. While the primary focus is adult speech, including children as contributors to the corpus can not only bring novelty to the project but will be a big step towards enabling the evaluation and validation of existing methods on children. Following the successful creation of the corpus, state-of-the-art machine learning and signal processing methods in speech emotion recognition can be evaluated for performance on the developed corpus, including investigating the methods for inherent biases [2] and other differences.

1.3 Hypothesis

Emotion classes and emotions states can be inferred from the speech signal of an individual up to a certain accuracy. Furthermore, the features responsible for speech emotion recognition are dependent on the language and age of the individual.

2 Study methods

2.1 Experimental design and measurement and post-processing

Experimental design: The experiment will be conduced in selected theatre groups/schools, potentially at different locations. We aim to create the dataset with similar sample distribution from adults and children. Therefore, up to 40 skilled actors will be recruited in the study. Of the 40 participants, we aim to have 20 adults, over the age of 18 and 20 children between 7-18 years of age. Actors with Danish skills and preferably also with English skills will be recruited. We select five sentences that will be enacted with five different emotions by all the participants. In order

to remove subjective associations and differences, we select the sentences such that they comprise of minimum contextual information. Each participant enacts all sentences for a specific emotion before moving on to the next emotion. Sentences in Danish are enacted first and if the participant agrees, they may also enact the sentences in English.

Measurement: The participants will be fitted with wireless Røde microphones which will be paired with their corresponding receiver. The recordings will be extracted from the receiver following the completion of the entire session by a participant.

Post-processing: Post recordings, the sentences from each actor will be segmented based on turns. Each turn is assumed to be a continuous flow of words without major pauses.

2.2 Tasks

The primary goal of the study is to explore the limits of inferring emotions from speech signals alone. Towards this end, the project is subdivided into three main tasks:

1. *Emotion representations:* How is emotion represented in speech signals and can emotion be disentangled from other factors influencing an individual's speech.
2. *Impact of age on emotion representations:* How does age impact the perception of emotions from speech and if age does have an impact, can the differences be modelled? As a supplement, which differences do we see between a population from 7-18 and a population above 18.
3. *Impact of language on emotion representations:* What influence does language have on emotion recognition from speech and what is the feasibility of an universal representation of emotions.

2.3 Analysis

Statistical analysis: We will use signal level descriptive signal statistics to investigate differences over emotion classes and age. The features include first and second order signal statistics and the statistics broadly corresponding to the tone, energy and spectral distribution of the signal [7]. Furthermore, we will explore the potential of modelling the emotions by mixed models where the subjects are included as random effects. The null hypothesis is that the above features are indistinguishable over (a) the emotion categories, (b) age. For the group differences, we shall conduct an analysis of variance (ANOVA) with 0.05 as significance level. The Benjamini-Hochberg False Discovery Rate (FDR) will be used to address false discoveries due to multiple hypothesis testing.

Clustering, preference and transfer learning: Unsupervised clustering, preference and transfer learning methods will be used to study the representations of the emotions over subjects. In addition to the aforementioned experiment data set, we will use public speech data sets for pre-training and transfer learning.

Software packages: Python (version>= 3.6), R (version>= 4), Matlab (version>= 2020*b*) and associated packages will be used for pre-processing, analysis and learning.

3 Participant information

3.1 Inclusion criteria

1. Part of an acting school, class, course or group; in other words, the individual should have had some acting training.
2. Age >= 7
3. Danish mandatory, English is a plus.

3.2 Exclusion criteria

1. Individuals who cannot speak Danish.
2. Age < 7

3.3 Reimbursement

Each participant will be reimbursed with 150 DKK, in the form of a gift card, for their participation in the study.

4 Ethical considerations

Written consent is obtained from the parent or their guardian before the experiment begins. If a participant withdraws during the experiment, the data will be deleted. However, the data will not be deleted after the experiment has finished.

4.1 Ethical implications

Since the speech sentences are pre-defined and selected to minimize subjective influences, and the emotions are enacted, the data collected does not comprise of any signals reflecting on the state of an individual and potential negative impact on the individuals is therefore minimized.

4.2 Data protection and handling

Before the experiment, the participant or their guardian will sign a consent form and fill a general questionnaire with questions on age and gender. The recorded speech signals and the participant information will be pseudo-anonymized by generating a random ID for each participant, and the randomly generated key will be accessible only to the main investigator of the project. The data and the general information will be stored on the secure storage at DTU and pseudo-anonymization of the data will be ensured by storing the consent with the identifying information and data separately without the possibility of tracing back to the individual IDs from the data.

5 Data sharing and dissemination of results

A public version of the data set will be created that can be used by researchers in the domain of speech emotion studies. This version will be pseudo-anonymized, comprising of speech signals only (i.e. not of gender and age), and will be shared with researchers who sign the end user licence agreement.

The findings from the study will be submitted for publication in a peer-reviewed journal and we anticipate at least one scientific paper resulting from the study. All publications will be performed such that the identity of the participants is protected.

6 Finances

- The initiator of the project is Line Katrine Harder Clemmensen, Ph.d., associate professor, DTU Compute.
- The project is funded by the larger WristAngel poject which is funded by an exploratory Synergy grant from the Novo Nordisk Foundation and is a collaboration with Copenhagen University Hospital, the Child Psychiatry Research Unit. The project received 668,107 euros.
- The research group has no financial conflicts of interest.

7 Study period (data collection only)

The data collection for the project will start in 04.2022 and end in 06.2023.

8 Future applications and implications of the proposed research

The project will establish a Danish speech data set comprising of emotionally rich speech samples from adults and children. The data set will enable the development of speech systems for the Danish speaking population. Furthermore, this corpus will also enable the validation of existing speech systems that were trained and developed for non-danish languages and mostly adults. Validation of such systems is essential as language, cultural influences, age and gender considerably influence the performance of speech systems and tasks.

References

[1] V. Petrushin, "Emotion in speech: Recognition and application to call centers," in *Proceedings of artificial neural networks in engineering*, vol. 710, p. 22, Citeseer, 1999.

[2] B. W. Schuller, "Speech emotion recognition: Two decades in a nutshell, benchmarks, and ongoing trends," *Communications of the ACM*, vol. 61, no. 5, pp. 90–99, 2018.

[3] C. Voss, J. Schwartz, J. Daniels, A. Kline, N. Haber, P. Washington, Q. Tariq, T. N. Robinson, M. Desai, J. M. Phillips, C. Feinstein, T. Winograd, and D. P. Wall, "Effect of Wearable Digital Intervention for Improving Socialization in Children With Autism Spectrum Disorder: A Randomized Clinical Trial," *JAMA Pediatrics*, vol. 173, pp. 446–454, 05 2019.

[4] P. Luana, "New kairos facial recognition camera offers customer insights," September 2019.

[5] I. S. Engberg, A. V. Hansen, O. Andersen, and P. Dalsgaard, "Design, recording and verification of a danish emotional speech database," in *Fifth European conference on speech communication and technology*, 1997.

[6] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. N. Chang, S. Lee, and S. S. Narayanan, "Iemocap: Interactive emotional dyadic motion capture database," *Language resources and evaluation*, vol. 42, no. 4, pp. 335–359, 2008.

[7] F. Eyben, K. R. Scherer, B. W. Schuller, J. Sundberg, E. André, C. Busso, L. Y. Devillers, J. Epps, P. Laukka, S. S. Narayanan, *et al.*, "The geneva minimalistic acoustic parameter set (gemaps) for voice research and affective computing," *IEEE transactions on affective computing*, vol. 7, no. 2, pp. 190–202, 2015.

Appendix I: Informationsbrev
Informationsbrev til deltageren (DK)

Kære deltager,
I dette brev oplyser vi dig om eksperimentets formål og fremgangsmåde. Det er vigtigt at du læser brevet omhyggeligt. Hvis du har nogle spørgsmål, skal du ikke tøve med at kontakte os (Line H. Clemmensen, lkhc@dtu.dk) for præciseringer.

Dine rettigheder som deltager
Din deltagelse i dette eksperiment er frivillig. Det betyder at du til enhver tid kan forlade eksperimentet uden at det har nogle konsekvenser for dig, uden at du behøver at give en begrundelse. Data, der er behandlet i forskningsprojektet inden du trækker dit samtykke tilbage, vil ikke blive slettet. Hvis du trækker dit samtykke tilbage midt i forsøget, vil vi slette dine data, og de kan dermed ikke behandles i forsknignsprojektet. Vi vil bede dig om en erklæring om informeret samtykke til at deltage, efter at du er blevet informeret om eksperimentet.

1 September 2021
LKHC

Formålet med forskningsprojektet
I dette forsøg undersøger vi tale- og stemme-signaler der fremføres under skuespil af forskellige følelser. Vi er især interesseret i at undersøge dansk tale, og dets forskelle til andre sprog samt stemme forskelle mellem børn og voksne. Forskningsresultaterne kan bl.a. bruges til at udvikle digitale assistenter til børn med psykiske lidelser som e.g. OCD (obsessive compulsive disorder).

Lagring og håndtering af data
Alt registreret forskningsdata (lydfiler, køn, alder) vil blive opbevaret pseudo-anonymiseret, det vil sige at det ikke vil have nogen forbindelse til dit navn, din adresse, dit CPR-nummer eller andre oplysninger, der gør det muligt at identificere dig. For at imødekomme dette opbevares din data under en nøglekode, som kun forskerne er bekendt med og som vil blive slettet, når den ikke er nødvendig længere (efter dataanalysen). Personlige oplysninger som dit navn vil blive opbevaret separat og lagret sikkert af forsøgslederen. En del af forskningsdataen (lydfiler og information om hvilken sætning der udtales med hvilken følelse under skuespil) vil blive gjort tilgængelig via DTUs datalager i pseudo-anonymiseret form med en tilfældig nøglekode, som ikke er gemt eller kendt af forskerne (køn og alder vil ikke blive gjort tilgængelige via DTUs datalager). Selv om identifikationsoplysninger er slettet er det muligt at nogen vil være i stand til at identificere dig i de offentliggjorte forskningsdata.

Følelsesgenkendelse via talesignaler
Vi vil benytte os af talesignaler udtalt under skuespil til at undersøge indikationer på følelser i stemmen. Lydoptagelser via mikrofon er en sikker og ikke-invasiv

DTU Compute
Department of Applied Mathematics and Computer Science
Richard Petersens Plads
Building 324
2800 Kgs. Lyngby
Denmark
Tel. +45 45 25 30 31
Fax +45 45 88 13 99
lkhc@dtu.dk
www.compute.dtu.dk

metode til at optage tale og stemme. I forsøget vil du blive bedt om at sige fem sætninger fem gange, hver gang med en ny følelse (f.eks. vrede, glæde, ked af det). Der vil være instruktioner og en forsøgsansvarlig til at hjælpe dig gennem forsøget.

Skulle du have spørgsmål til eksperimentet, de anvendte metoder eller din egen sikkerhed samt dine rettigheder, så tøv ikke med at kontakte os per e-mail (lkhc@dtu.dk) eller telefonisk (+45 45 25 xx xx).
Med venlig hilsen, Line H. Clemmensen, Lektor

DTU privacy rules:
https://www.dtu.dk/english/About/strategy-policy/Policies/Privacy-policy
https://www.dtu.dk/Om-DTU/strategi_aarsrapporter_mv/Politikker/Privatlivspolitik

Appendix II: Informeret samtykke

Erklæring om informeret samtykke (DK)

Eksperiment: DanskEmoTale (DanishEmoSpeech)
Forsker: Line K. H. Clemmensen, Lektor

Erklæring om informeret samtykke

1 September 2021
LKHC

Jeg bekræfter at:
- Jeg er blevet tilfredsstillende informeret om det pågældende forsøg, både mundtligt og skriftligt i form af informationsbrevet.
- Jeg har haft muligheden for at fremlægge spørgsmål vedrørende forsøget og at disse spørgsmål er blevet besvaret på tilfredsstillende vis.
- Jeg har overvejet min deltagelse i eksperimentet omhyggeligt.
- Jeg deltager af min egen fri vilje.

Jeg er enig i at:
- Mine data vil blive indsamlet og opbevaret til videnskabelige formal som nævnt i informationsbrevet i pseudo-anonymiseret tilstand, dvs. med en nøglekode som kun forskerne er bekendt med.
- Mine data må blive delt i en offentlig forskningsdatabase i pseudo-anonymiseret tilstand, med en tilfældig nøglekode, der ikke er gemt af forskerne. Dette inkluderer lydfiler samt labels med den følelse der er udøvet under skuespil i lydfilerne. Navn, alder og køn er IKKE inkluderet.

Jeg forstår at:
- Min deltagelse er frivillig og jeg har ret til at trække mig fra eksperimentet til enhver tid uden at give en begrundelse.
- Hvis jeg tilbagekalder min deltagelse i forsøget, må mine forskningsdata stadigvæk anvendes til forskningsformål.
- Mit privatliv er beskyttet i henhold til dansk lovgivning og europæiske retningslinjer (GDPR; EU 2016/679).
- Der er en risiko for, at nogen muligvis kan identificere mig i offentligt delte forskningsdata.

DTU Compute
Department of Applied Mathematics and Computer Science
Richard Petersens Plads
Building 324
2800 Kgs. Lyngby
Denmark
Tel. +45 45 25 30 31
Fax +45 45 88 13 99
lkhc@dtu.dk
www.compute.dtu.dk

Jeg giver mit samtykke til at deltage i dette eksperiment

_____

_____

Fulde navn Fødselsdato

_____

_____

Dato, Place Signature

Appendix IIb: Informeret samtykke, barn

Erklæring om informeret samtykke (DK)

Eksperiment: DanskEmoTale (DanishEmoSpeech)
Forsker: Line K. H. Clemmensen, Lektor

Erklæring om informeret samtykke (barn)

1 September 2021
LKHC

Jeg bekræfter at:
- Mit barn og jeg er blevet tilfredsstillende informeret om det pågældende forsøg, både mundtligt og skriftligt i form af informationsbrevet.
- Mit barn og jeg har haft muligheden for at fremlægge spørgsmål vedrørende forsøget og at disse spørgsmål er blevet besvaret på tilfredsstillende vis.
- Mit barn og jeg har overvejet mit barns deltagelse i eksperimentet omhyggeligt.
- Mit barn deltager af egen fri vilje.

Jeg er enig i at:
- Mit barns data vil blive indsamlet og opbevaret til videnskabelige formal som nævnt i informationsbrevet i pseudo-anonymiseret tilstand, dvs. med en nøglekode som kun forskerne er bekendt med.
- Mit barns data må blive delt i en offentlig forskningsdatabase i pseudo-anonymiseret tilstand, med en tilfældig nøglekode, der ikke er gemt af forskerne. Dette inkluderer lydfiler samt labels med den følelse der er udøvet under skuespil i lydfilerne. Navn, alder og køn er IKKE inkluderet.

Jeg forstår at:
- Mit barns deltagelse er frivillig og jeg har ret til at trække mit barn fra eksperimentet til enhver tid uden at give en begrundelse.
- Hvis jeg tilbagekalder mit barns deltagelse i forsøget, må mit barns forskningsdata stadigvæk anvendes til forskningsformål.
- Mit barns privatliv er beskyttet i henhold til dansk lovgivning og europæiske retningslinjer (GDPR; EU 2016/679).
- Der er en risiko for, at nogen muligvis kan identificere mit barn i offentligt delte forskningsdata.

DTU Compute
Department of Applied Mathematics and Computer Science
Richard Petersens Plads
Building 324
2800 Kgs. Lyngby
Denmark
Tel. +45 45 25 30 31
Fax +45 45 88 13 99
lkhc@dtu.dk
www.compute.dtu.dk

Jeg giver mit samtykke til at mit barn må deltage i dette eksperiment

_____

_____
Fulde navn, Fødselsdato på barn

_____

_____
Fulde navn, Fødselsdato på forælder/værge

_____

_____
Dato, sted, underskrift forælder/værge

Appendix III

INSTRUCTIONS

1) You will be given a copy of the consent form. Please read it carefully and let us know if you disagree with any of the terms.
2) If you agree with it, kindly sign the consent~(add your name below your signature).
3) A recording-session constitutes of 5 sets small sessions; in each session you will have to enact the 5 sentences provided to you with a single emotion.
4) Before the start of every recording, you will be fitted with a collar microphone.
5) Once the recording sessions begin, enact the 5 sentences with the requested emotion (eg: anger, happy, sad) one after another, with distinguishable pauses between the sentences.
6) After completing each emotion, move on to the next emotion in the list and repeat step 5.
7) Feel free to ask in case of any questions.

INSTRUKTIONER

1. Du får udleveret en kopi af samtykkeerklæringen. Læs den omhyggeligt, og lad os vide, hvis du er uenig i nogen af vilkårene.
2. Hvis du er enig i erklæringen, skal du venligst underskrive samtykket~(tilføj dit navn under din underskrift).
3. En optagelsessession består af 5 små sessioner; i hver session skal du udføre de 5 sætninger du får med en og samme følelse.
4. Før starten af hver optagelse vil du blive udstyret med en kravemikrofon.
5. Når optagelsessessionerne begynder, skal du udføre de 5 sætninger med den ønskede følelse (f.eks.: vrede, glæde, ked af det) efter hinanden, med tydelige pauser mellem sætningerne.
6. Efter at have fuldført hver følelse, skal du gå videre til den næste følelse på listen og gentage trin 5.
7. Du er velkommen til at spørge, hvis du har spørgsmål.