Masking speech contents by random splicing: Is emotional expression preserved?

F. Burkhardt[1], Anna Derington[1], Matthias Kahlau[1], Klaus Scherer[2], F. Eyben[1], B.W. Schuller[1,3,4]
[1]audEERING GmbH, [2]University of Geneva, Switzerland, [3]University of Augsburg, [4]Imperial College London

Introduction & Summary

- We discuss the influence of random splicing on the perception of emotional expression in speech signals.
- Use cases:

(a) anonymizing the speech for privacy protection, mainly be removing personal information (used in ECoWEB)

(b) removing the linguistics to force human annotators of emotional expression to focus on the extra linguistic features and not on the linguistic content, thereby enabling training annotation that is valid across languages

Random splicing

In two steps, now part of openSMILE framework (component called "AudioScrambler")

Segmentation

Segment length not fixed, but is determined by configuring a minimum and maximum fragment length. Within which the root-mean-square (RMS) values in the time domain used to determine best cutting point.

Rearrangement

In the second step, the resulting segments are rearranged in a pseudo-random order by shuffling their list indices, in such a way that no segment is connected to any segment that was already connected there before, unless it is the last remaining segment.

The dataset

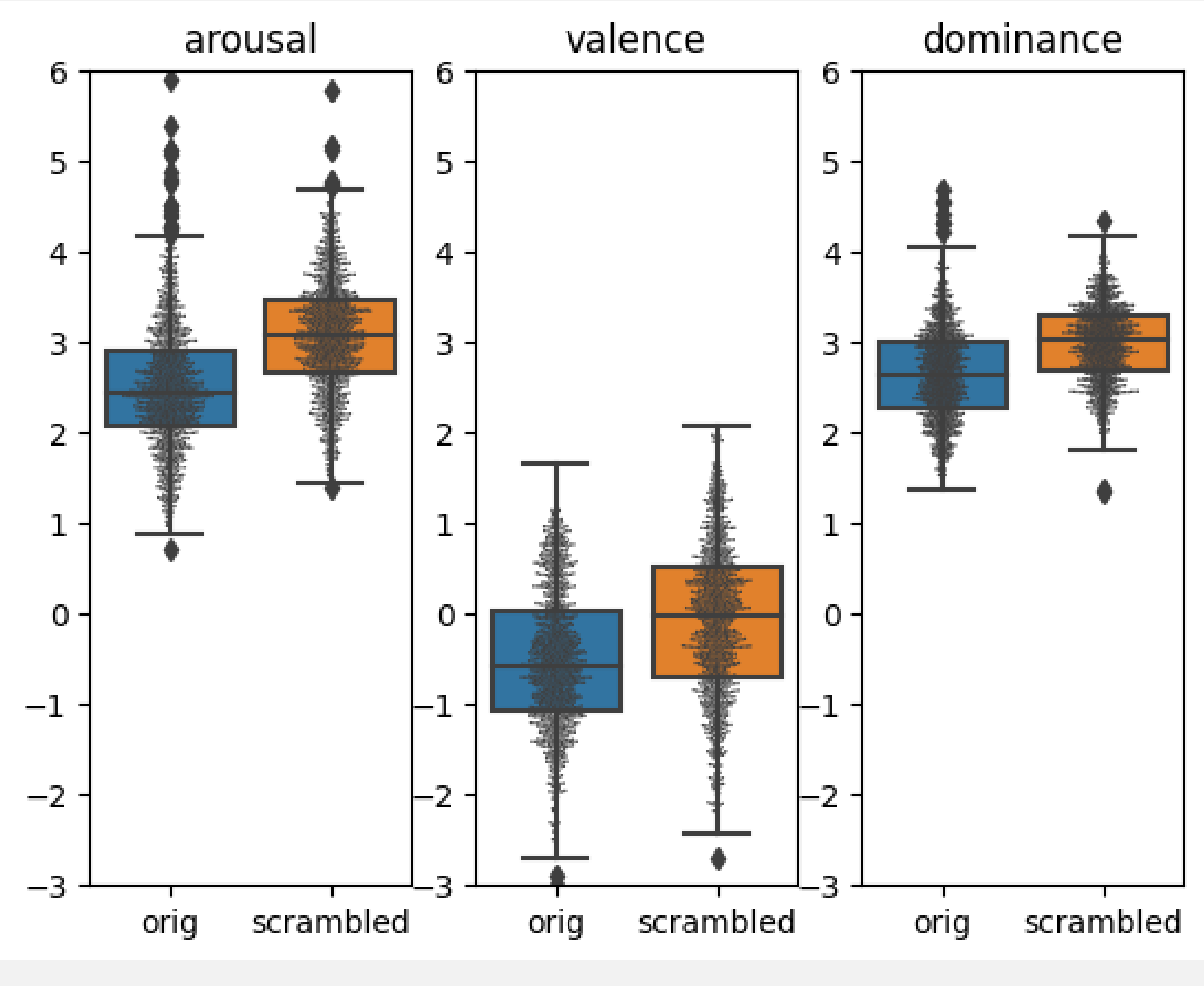
German Parliament samples

We tested the approach on a database of German parliament speeches. The database contains data from 9 German politicians. After a manual segmentation the data consists of 1198 segments spoken by the nine politicians. The age span was from 40 to 77 years, with 6 men and 3 women. For reproducibility, the data can be accessed via Zenodo https://zenodo.org/record/7224678

Manual Evaluation

12 annotators employed by audEERING GmbH rated the whole set of original segments with respect to the three dimensions *arousal*, *valence* and *dominance*.

In the Figure, we depict the distributions of the labels for arousal, valence and dominance for random-spliced and original samples, respectively. It can be easily seen that the labels differentiate mainly for the arousal dimension. The majority of the valence labels are negative which is probably due to the domain: politicians speaking in parliament. Arousal and dominance are both clearly on the positive side, which also makes sense with respect to the domain. It seems that arousal also gets overestimated for the random spliced versions.



Manual Evaluation

Although all T-tests clearly show that the influence of random splicing on the samples is a highly significant one, the correlation between the corresponding samples is quite high, especially for arousal and least for valence, which tends to be over-estimated for the random spliced samples.

	PCC	CCC	pairwise t-test
Arousal	.785	.548	> .001
Valence	.524	.519	> .001
Dominance	.603	.545	> .001

Table: Results of the manual annotations

Model Evaluation

We computed arousal/valence/dominance predictions with a pre-trained model: Wav2vec2.0 finetuned on MSPPodcast.

	CCC orig	CCC scrambled
Arousal	.409	.405
Valence	.080	.135
Dominance	.351	.319

Table: Results of the model predictions

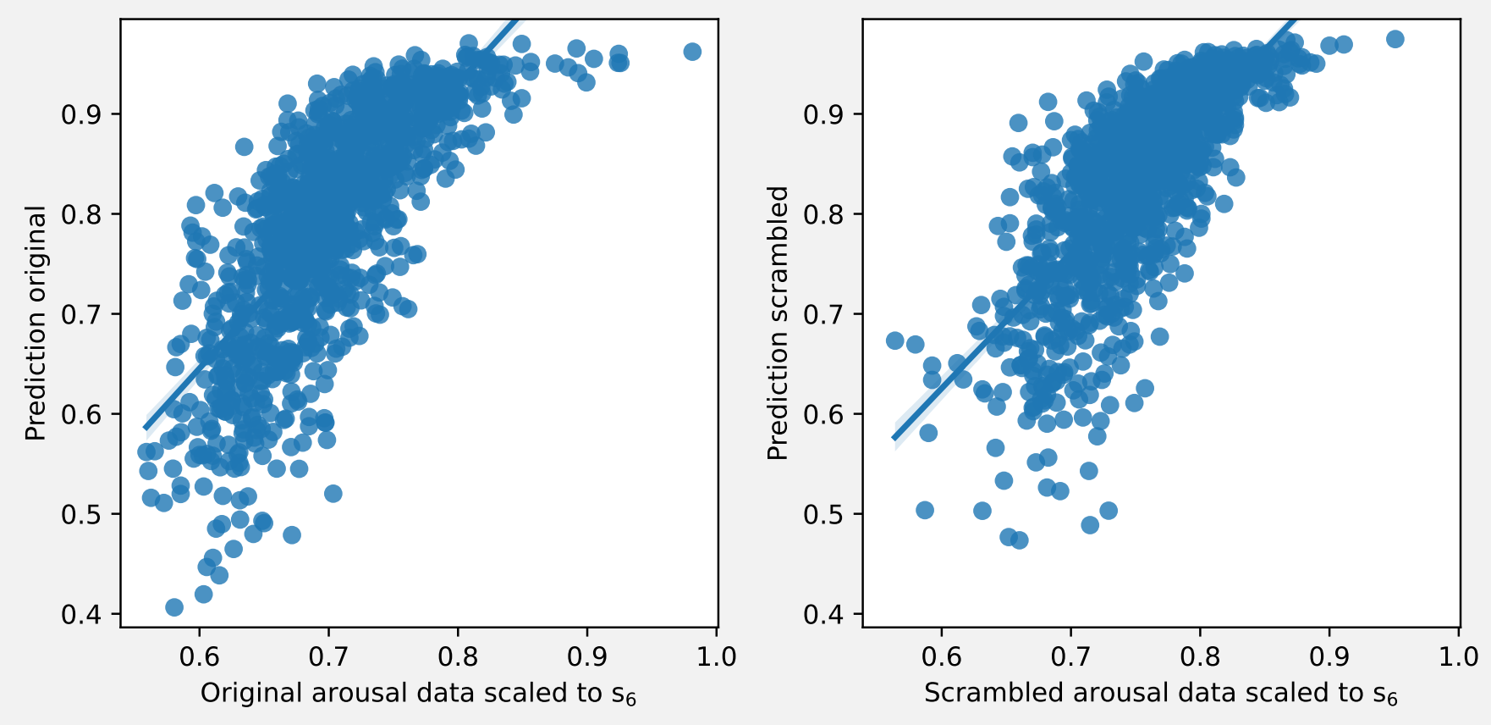


Figure: Original and scrambled model predictions correlation for arousal

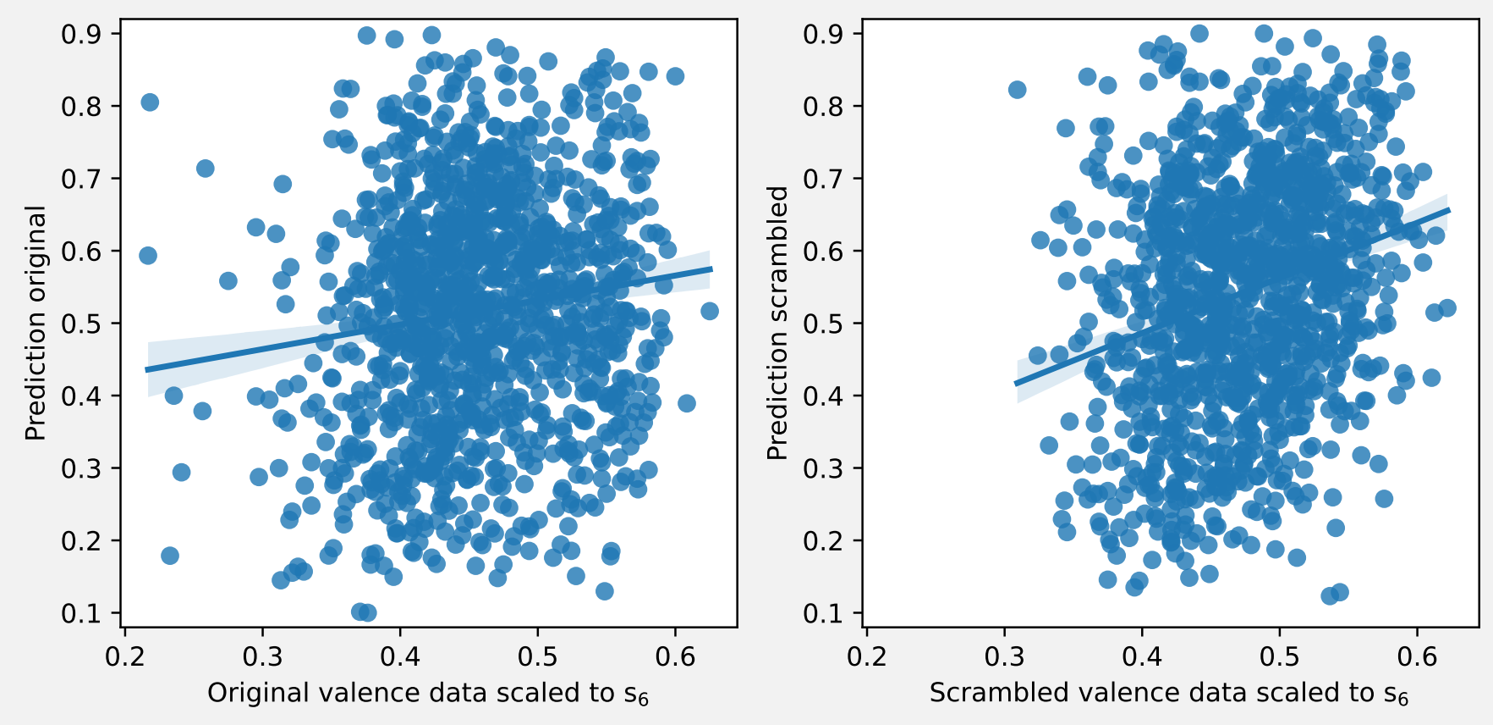


Figure: Original and scrambled model predictions correlation for valence

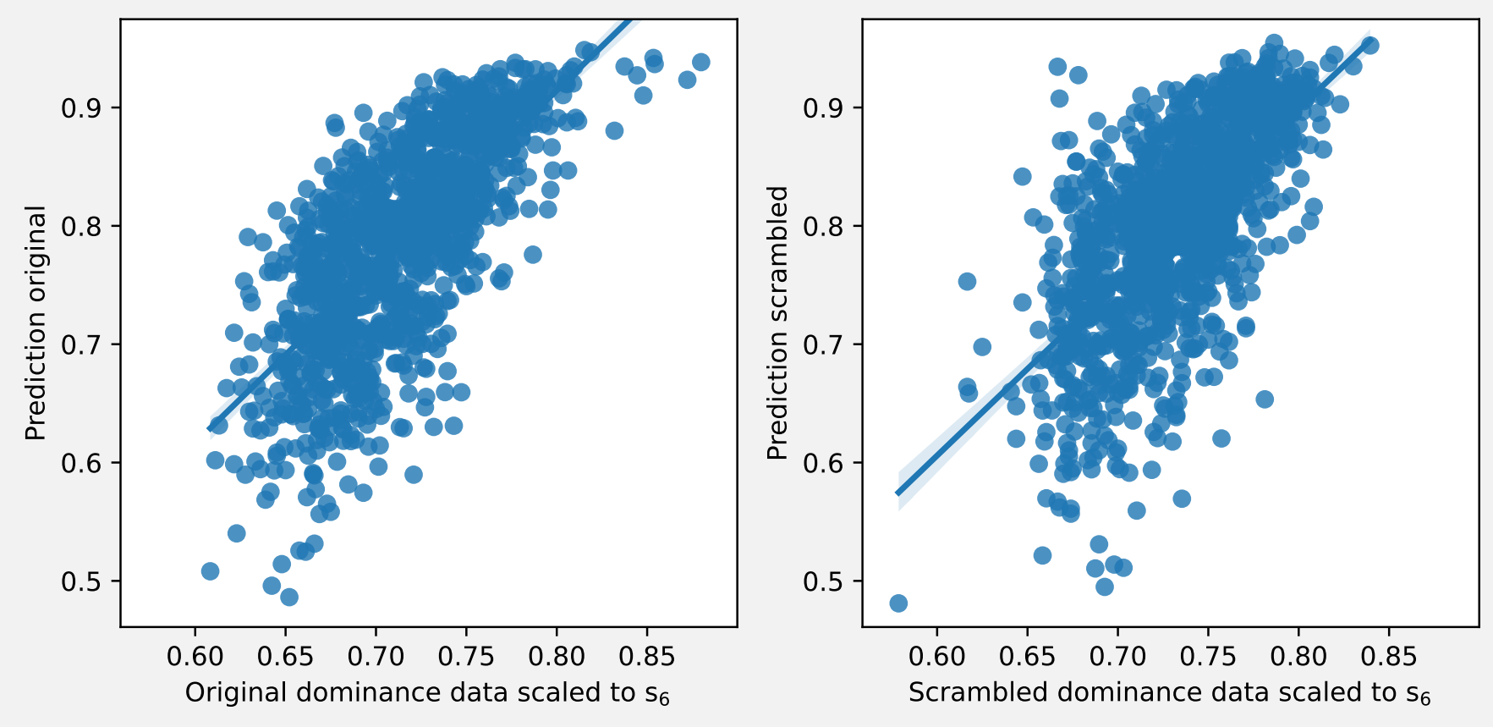


Figure: Original and scrambled model predictions correlation for dominance

- Wav2vec2.0 models have been shown to exploit linguistic data, the evaluation model has been trained on English data only.
- Therefore, it is understandable that the model performs better on the valence task, where the linguistic component has been removed.

Outlook

- Future investigations could deal with more elaborate splicing algorithms which may be informed by linguistic embeddings in order to less disrupt valence aspects.

Acknowledgements

This research has been partly funded by the European ECoWeB (Assessing and Enhancing Emotional Competence for Well-Being in the Young) project (Grant Agreement number: 754657) as well as the European MARVEL (Multimodal Extreme Scale Data Analytics for Smart Cities Environments) project (Grant Agreement ID: 975337.