Speech Emotion Recognition among Couples using the Peak-End Rule and Transfer Learning

George Boateng
gboateng@ethz.ch
ETH Zürich
Zurich, Switzerland

Laura Sels
laura.sels@ugent.be
Ghent University
Ghent, Belgium

Peter Kuppens
peter.kuppens@kuleuven.be
KU Leuven
Leuven, Belgium

Peter Hilpert
p.hilpert@surrey.ac.uk
University of Surrey
Surrey, United Kingdom

Tobias Kowatsch
tkowatsch@ethz.ch
ETH Zürich
Zurich, Switzerland
University of St. Gallen
St. Gallen, Switzerland

ABSTRACT

Extensive couples' literature shows that how couples feel after a conflict is predicted by certain emotional aspects of that conversation. Understanding the emotions of couples leads to a better understanding of partners' mental well-being and consequently their relationships. Hence, automatic emotion recognition among couples could potentially guide interventions to help couples improve their emotional well-being and their relationships. It has been shown that people's global emotional judgment after an experience is strongly influenced by the emotional extremes and ending of that experience, known as the peak-end rule. In this work, we leveraged this theory and used machine learning to investigate, which audio segments can be used to best predict the end-of-conversation emotions of couples. We used speech data collected from 101 Dutch-speaking couples in Belgium who engaged in 10-minute long conversations in the lab. We extracted acoustic features from (1) the audio segments with the most extreme positive and negative ratings, and (2) the ending of the audio. We used transfer learning in which we extracted these acoustic features with a pre-trained convolutional neural network (YAMNet). We then used these features to train machine learning models — support vector machines — to predict the end-of-conversation valence ratings (positive vs negative) of each partner. The results of this work could inform how to best recognize the emotions of couples after conversation-sessions and eventually, lead to a better understanding of couples' relationships either in therapy or in everyday life.

CCS CONCEPTS

• **Applied computing → Psychology**.

KEYWORDS

Speech emotion recognition; Speech processing; Affective computing; Couples; Transfer Learning; Peak-end rule; Convolutional neural network; Support vector machine

ACM Reference Format:
George Boateng, Laura Sels, Peter Kuppens, Peter Hilpert, and Tobias Kowatsch. 2020. Speech Emotion Recognition among Couples using the Peak-End Rule and Transfer Learning. In *Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion), October 25–29, 2020, Virtual event, Netherlands.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3395035.3425253

1 INTRODUCTION

Couples' observation research has shown that the emotions that couples experience during a conflict predict if these couples stay together in the long-term (for an overview, see [19]). For instance, couples heading for break-up show more negative emotions and less positive emotions than happy couples, and are stuck in certain emotional patterns [7, 18]. Although couples' observation research has delivered valuable clinical insights, it also suffers from measurement issues such as low cross-validity and interrater reliability [23] and entails some methodological challenges. One important methodological challenge is the manual coding of audio-video data, which is very costly and time-consuming [27]. Automated emotion recognition could alleviate these limitations, and therefore advance the field in important ways [36].

Several emotion recognition works on couple dyads use data that is collected from individuals acting out dyadic interactions either using a script or engaging in spontaneous sessions [5, 6, 32, 34]. A lot of emotion recognition works use such data sets [38]. The emotions are later rated by others amidst several challenges [33] and do not necessarily reflect the subjective emotions of the individuals. Additionally, these algorithms are likely to perform poorly on naturalistic data [10].

On the other hand, there are few works on detecting the emotional behavior of real couples. Some leveraged interaction dynamics among the partners (e.g. entrainment — synchrony between partners) [2, 28, 29] and salient instances [16, 17, 26] to perform recognition. These works tend to use emotion labels from external

ACM ISBN 978-1-4503-8002-7/20/10…$15.00
https://doi.org/10.1145/3395035.3425253

raters rather than the couples and hence do not reflect the subjective emotions of the couples.

Our aim is to build upon recent findings from fundamental psychological research to automatically recognize couples' self-reported emotions. Specifically, couples literature has shown that how couples feel after a conflict is predicted by certain emotional aspects of that conversation (e.g., [13, 14, 21, 30, 31]); and recently, it has been suggested that the emotional extremes and ending of the conversation might be particularly valuable [44]. In fact, in a variety of domains, it has been shown that judgments of emotional experiences are most impacted by the most extreme moments (peaks) and the end of that particular experience, known as the peak-end rule [12, 25]. The peak-end rule could be leveraged to develop systems to better recognize the emotions of couples.

Building upon our recommendations in [4], we investigate through a machine learning perspective which segment(s) of an audio conversation could be used to best recognize the emotions of each partner after a conversation. Our research question is as follows:

Using features of which of the following audio segments produce the best emotion recognition result: a) segments with the most extreme positive and negative ratings, b) the ending of the audio or c) a combination of the extremes and ending?

In this first of its kind work, our primary contribution is the exploration of the best way to recognize the emotions of couples after a conversation (5 - 10 minutes) through the peak-end rule lens using deep learning approaches. Our secondary contribution is the use of a unique dataset — real-world data collected from Dutch-speaking couples with self-ratings of emotions. Our third contribution is our proposal and computation of a "partner perception baseline" for emotion recognition within the context of couples interactions that leverage each partner's perception of his/her partner's emotions.

We classified the end-of-conversation valence (positive vs negative) of Dutch-speaking couples using acoustic features from various segments of the audio and compared with the partner perception baseline. We used transfer learning, an approach used in deep learning to circumvent the need to develop hand-crafted features [11]. It is used to address the limitations of using small labeled datasets and has shown success in various fields including emotion recognition tasks ([35, 42]). The results of this work would inform the best way to recognize the emotions of couples' after conversation-sessions and eventually, lead to a better understanding of couples' relationships either in therapy or in everyday life.

The rest of this paper is organized as follows: In Section 2, we describe our method. In Section 3, we describe our experiments In Section 4, we show and discuss the results. In Section 5, we present limitations of this work and future work, and we conclude in Section 6.

2 METHODS

In this section, we describe the dataset and preprocessing, and the transfer learning approach (Figure 1).

2.1 Dataset and Preprocessing

A Dyadic Interaction lab study was conducted in Belgium with 101 Dutch-speaking, heterosexual couples. These couples were first asked to have a 10-minute conversation about a negative topic (a characteristic of their partner that annoys them the most), followed by a 10-minute conversation about a positive topic (a characteristic of their partner that they value the most) [9, 43–45]. During both conversations, couples were asked to wrap up the conversation after 8 minutes. For the negative topic, they were also asked to end on good terms. After each conversation, each partner watched the video recording of the conversation separately on a computer and rated his or her emotion on a moment-by-moment basis by continuously adjusting a joystick to the left (very negative) and the right (very positive), so that it closely matched their feelings, resulting in valence scores on a continuous scale from -1 to 1 [20, 39]. Additionally, each partner reported how they felt after the interaction and also what they thought their partner felt, using the Affect Grid questionnaire [41]. The Affect Grid captures the valence and arousal dimensions of Russell's circumplex model of emotions [40].

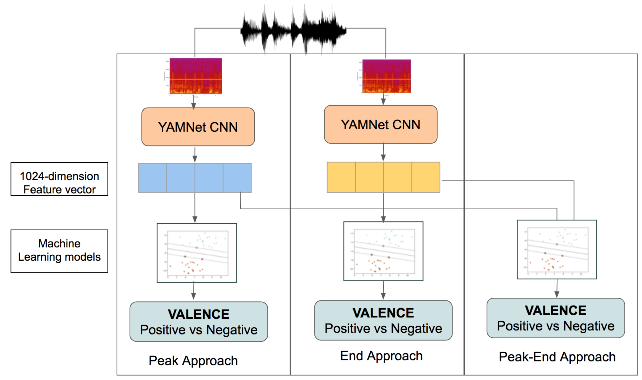

Figure 1: Overview of Approach

Valence refers to how negative to positive the person feels and arousal refers to how sleepy to active a person feels. Using these two dimensions, categorical emotions can be placed and grouped into the four quadrants: high arousal and negative valence (e.g. stressed), low arousal and negative valence (e.g. depressed), low arousal and positive valence (e.g. relaxed) and high arousal and positive valence (e.g. excited). Subjects had to place an 'x' on any square on the Affect Grid corresponding to their feelings about each conversation, which translates to a value of between 0 and 8 each for pleasure and arousal. We only used the valence dimension of the Affect Grid because the continuous rating that the end-of-conversation emotion was compared with was done only using valence. The continuous rating was restricted to valence to minimize the time spent by subjects in the lab and also because it is standard practice in such dyadic interaction designs. We categorized the valence scores into two classes, negative (0-4) and positive valence (5-8) for males and females. Also, we only used audios from the negative/conflict conversation in this work. We could use only 92 out of the 101 audios in this work as some of the data was unavailable due to several issues peculiar of real-world data collection such as missing self-ratings due to failure of the recording device, lack of speaker annotations for all couples among others. In total, for males, we had 22 negative and 70 positive ratings and for females, we had 16 negative and 76 positive ratings. This distribution shows how significantly imbalanced the dataset is which is reflective of real-world data and consistent with other couple emotion recognition works (e.g. [8]).

The audio was manually annotated showing which partner was speaking at various points of the audio. Trained research assistants (5) were instructed to listen and visually inspect the audios, and annotate the exact start and end of each talking turn for each partner. In addition, students coded pauses, cross-talk, and noise and laughter. Multiple rounds of checking were done to ensure this process was precisely done. We used the segments of the audio where the male or female spoke to extract audio segments corresponding to the peaks and ends for each partner. For the peaks, we used the continuous valence rating to find the specific second with the largest negative value (minimum) and the specific second with the largest positive value (maximum). We then used the speaker turn containing that specific second as the peak segment (each for the minimum and maximum). The average duration of the peak segments for all the couples was 3.5 seconds. For the ending, we used the last 60 seconds of the audio corresponding to 10% of the whole audio (600 seconds). There was no reference in the literature for the duration to use for the end and so we picked 60 secs (the last 10%) as we reasoned it will capture a good enough duration of each couple's interaction without being too long.

Finally, we computed a partner perception baseline for the context of emotion recognition among couples. We used the assessment of each partner's perception of his/her partner's emotion at the end of the conversation to compute the baseline. This baseline gives an estimate of how well each partner could infer his/her partner's emotion after an interaction. We argue this is a good enough human baseline with which to compare the machine learning approach since a person's partner, in theory, is the best person to know him or her albeit this perception is biased in practice [46].

2.2 Transfer Learning Approach

Given that the data set is small, we sought to leverage work that has been done for a related task and hence used transfer learning [35] where we used a model that is pre-trained on a similar problem. We extracted spectrograms and used a pretrained convolutional neural network (CNN) to compute embeddings as acoustic features which we used to perform classification with machine learning models. We used the YAMNet model [1] which is a CNN that was pretrained on the AudioSet dataset to predict 521 audio event classes [15, 22]. YAMNet is based on the MobileNet architecture [24]. We used the YAMNet model as a feature extractor and hence replaced the original final logistic layer which outputs 521 class with a linear support vector machine (SVM) which we trained.

We extracted a spectrogram as an input into the YAMNet model in the same way as was done for the trained model. The audio's sample rate is 16 kHz. A spectrogram is computed using magnitudes of the Short-Time Fourier Transform with a window size of 25 ms, a window hop of 10 ms, and a periodic Hann window. A mel spectrogram is computed by mapping the spectrogram to 64 mel bins covering the range 125-7500 Hz. A stabilized log mel spectrogram is computed by applying log(mel-spectrum + 0.01) where the offset is used to avoid taking a logarithm of zero. These features are then framed into non-overlapping examples of 0.96 seconds, where each example covers 64 mel bands and 96 frames of 10 ms each [1]. This resulted in a 2D data of size 96 x 64 for each second, which we used as a data point input to the YAMNet model. The output of the model is a 1024-dimensional feature vector per data point input of size 96 x 64. We then normalized the vectors to be zero mean and unit variance and then used the features vectors as inputs to a linear SVM.

Table 1: Results for Peak, End and Peak-End Approaches and Baseline

Approach	Balanced Accuracy (%)	
	Male	Female
Partner perception	73.2	74.3
Peak	48.8	74.8
End	50	58.6
Peak-End	53.3	54.4

3 EXPERIMENTS

We performed various experiments using a linear SVM and the scikit-learn library [37]. We trained models separately for males and females to perform binary classification of valence. Our main models were trained using features from the peak, end, and peak and end (peak-end). We used majority voting of the classification for all features to decide the class for the audio segment. We performed an evaluation with leave-one-couple-out cross-validation similar to [8] which is a robust evaluation approach and gives an estimate of how well the model will perform on an unseen couple. We used confusion matrices and the metric balanced accuracy for evaluation since the data is imbalanced. Balanced accuracy is the unweighted average of the recall of each class. We used different values of the hyperparameter "C" ranging from 10^{-4} to 10^{1} for separate models and present results for the hyperparameter that produced the best results. We used the "balanced" hyperparameter for all models of the SVM to account for the class imbalance while training. We compared our results to a random baseline equivalent to 50% balanced accuracy and our proposed partner perception baseline.

4 RESULTS AND DISCUSSION

We report the results of the best performing models in Table 1. The peak approach which used about only 1.1% of the whole 10 minute audio performed the best for the female model with 74.8%, outperforming both the random and partner perception baselines. Yet, it performed the worst for the male model. The peak-end approach performed the best for the male model with 53.3% albeit worse than the partner perception baseline and slightly better than the random baseline. Figure 2 and 3 show the confusion matrices of the best models for male and female respectively.

The peaks performing better than the end in predicting end-of-conversation affect (though for female partners only) is consistent with the results of [44], in which the peak rating was more predictive than the end. The peak approach produced the best results likely because the peak segments contained the most extreme emotional expressions (acoustically). This result was not the same for the male partners for whom the results for peak and ends were similar and

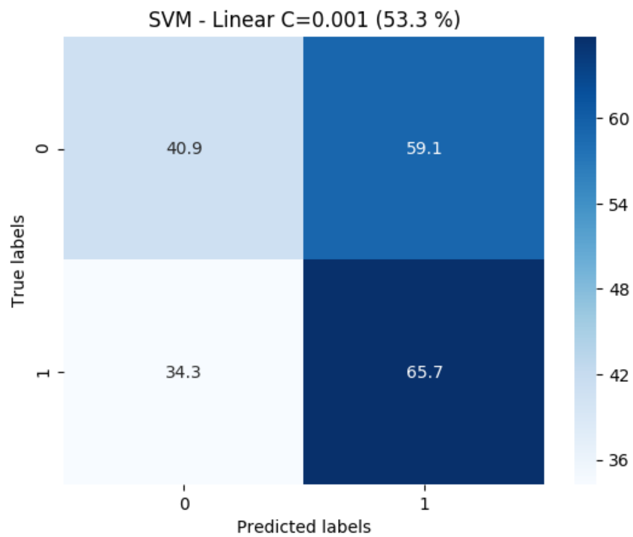

Figure 2: Best Male Result Confusion Matrix (Peak-End Approach)

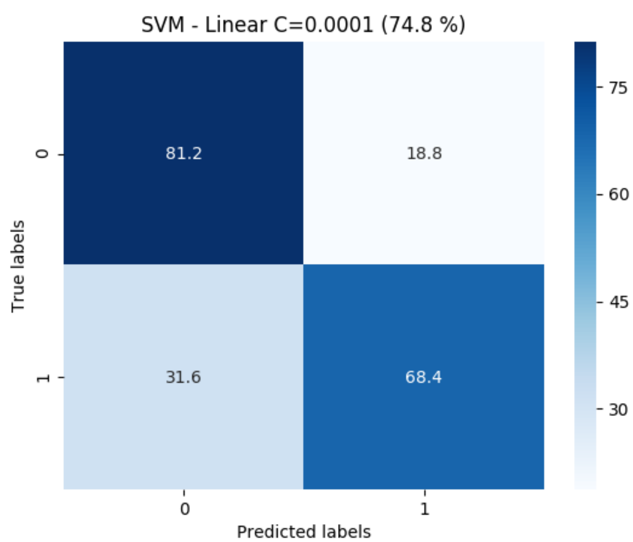

Figure 3: Best Female Result Confusion Matrix (Peak Approach)

worse than the baselines. This result suggests that the male partners may not have been more emotionally expressive (acoustically) at the peak segments than at the end. This reasoning is speculative and hence further investigation is needed using, for example, linguistic features before any conclusions can be drawn. These results points to the need to develop methods that can automatically identify the speaker turns with the most extreme emotional expressions, after which acoustic features can be extracted to get accurate end-of-conversation emotion predictions. This work is one step towards our goal to recognize the emotions of German-speaking couples in daily life based on 5 minutes of multimodal data from conversation moments which we are currently collecting [3].

5 LIMITATIONS AND FUTURE WORK

In this work, we did not perform an evaluation with the whole audio or random segments as the focus was on the peaks and ends. Hence, we used random and partner perception baselines for comparison. Future work will use the whole audio, and random segments. Also, we focused on valence since that was the only dimension rated in the continuous rating. Future work will need to collect data with the arousal dimension and explore using the arousal dimension. Those results could be used together with this work to identify the right quadrant of the Affect grid and consequently, the kinds of emotions the person may be feeling. Additionally, we only used the negative/conflict conversation. These experiments will be repeated with the positive conversation and results will be compared to the results of this work. Furthermore, this work focused on evaluating the segments using acoustic features. We currently do not have manual transcripts of the data and automatic speech recognition systems that we tried out did not work well for this Dutch-based speech data. Hence, we plan to get manual transcript of this data and use linguistic features also. Additionally, given that the continuous ratings were done for the whole conversation including the speech of both partners, the peak rating of each partner may not always overlap with a speech segment of that partner. Hence, we first extracted the speaker turns of each partner, and then found the speaker turn with the peak rating. Consequently, the most extreme rating overall may not have used. We extracted and used features from both positive and negative peaks. Future will evaluate using the positive and negative peaks separately and using different durations surrounding the peaks and ends. Additionally, we plan to perform a similar evaluation using self-reports other than the Affect Grid such as ratings for happy, sad, etc.

6 CONCLUSION

In this work, we performed an evaluation of the segments of an audio conversation that best predicts the end-of-conversation emotions of couples. We leveraged the peak-end rule, and a used transfer learning approach to extract features from (1) the audio segments with the most extreme positive and negative ratings, and (2) the ending of the audio. We used a pre-trained CNN to extract these acoustic features and a linear SVM to perform binary classification of the valence of partners. Our results showed that the segments from the peak produce the best results for recognizing the emotions of female partners and the approach was better than the partner perception baseline. This first-of-its-kind work contributes an evaluation of an approach that could be leveraged to best recognize the emotions of couples and then potentially used to improve the emotional well-being and relationship quality of couples via interventions.

ACKNOWLEDGMENTS

We are grateful to Markus Wyss, Arthur Deschamps and Daniel Végh for their contributions to this work.

REFERENCES

[1] [n.d.]. YAMNet. https://github.com/tensorflow/models/tree/master/research/audioset/yamnet.

[2] Matthew P Black, Athanasios Katsamanis, Brian R Baucom, Chi-Chun Lee, Adam C Lammert, Andrew Christensen, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2013. Toward automating a human behavioral coding system for married couples' interactions using speech acoustic features. *Speech communication* 55, 1 (2013), 1–21.

[3] George Boateng. 2020. Towards Real-Time Multimodal Emotion Recognition among Couples. In *Proceedings of the 2020 International Conference on Multimodal Interaction (ICMI '20), October 25–29, 2020, Virtual event, Netherlands.*

[4] George Boateng, Laura Sels, Peter Kuppens, Janina Lüscher, Urte Scholz, and Tobias Kowatsch. 2020. Emotion Elicitation and Capture among Real Couples in the Lab. In *1st Momentary Emotion Elicitation & Capture workshop (MEEC 2020), co-located with the ACM CHI Conference on Human Factors in Computing Systems.*

[5] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N Chang, Sungbok Lee, and Shrikanth S Narayanan. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. *Language resources and evaluation* 42, 4 (2008), 335.

[6] Carlos Busso, Srinivas Parthasarathy, Alec Burmania, Mohammed AbdelWahab, Najmeh Sadoughi, and Emily Mower Provost. 2016. MSP-IMPROV: An acted corpus of dyadic interactions to study emotion perception. *IEEE Transactions on Affective Computing* 8, 1 (2016), 67–80.

[7] Laura L Carstensen, John M Gottman, and Robert W Levenson. 1995. Emotional behavior in long-term marriage. *Psychology and aging* 10, 1 (1995), 140.

[8] Sandeep Nallan Chakravarthula, Haoqi Li, Shao-Yen Tseng, Maija Reblin, and Panayiotis Georgiou. 2019. Predicting Behavior in Cancer-Afflicted Patient and Spouse Interactions Using Speech and Language. *Proc. Interspeech 2019* (2019), 3073–3077.

[9] Egon Dejonckheere, Merijn Mestdagh, Marlies Houben, Isa Rutten, Laura Sels, Peter Kuppens, and Francis Tuerlinckx. 2019. Complex affect dynamics add limited information to the prediction of psychological well-being. *Nature human behaviour* 3, 5 (2019), 478–491.

[10] Sidney K D'mello and Jacqueline Kory. 2015. A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR)* 47, 3 (2015), 1–36.

[11] Kexin Feng and Theodora Chaspari. 2020. A Review of Generalizable Transfer Learning in Automatic Emotion Recognition. *Frontiers in Computer Science* 2 (2020), 9.

[12] Barbara L Fredrickson. 2000. Extracting meaning from past affective experiences: The importance of peaks, ends, and specific emotions. *Cognition & Emotion* 14, 4 (2000), 577–606.

[13] Lisa Gaelick, Galen V Bodenhausen, and Robert S Wyer. 1985. Emotional communication in close relationships. *Journal of personality and social psychology* 49, 5 (1985), 1246.

[14] Robert L Geist and David G Gilbert. 1996. Correlates of expressed and felt emotion during marital conflict: Satisfaction, personality, process, and outcome. *Personality and Individual Differences* 21, 1 (1996), 49–60.

[15] Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio set: An ontology and human-labeled dataset for audio events. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 776–780.

[16] James Gibson, Athanasios Katsamanis, Matthew P Black, and Shrikanth Narayanan. 2011. Automatic identification of salient acoustic instances in couples' behavioral interactions using diverse density support vector machines. In *Twelfth Annual Conference of the International Speech Communication Association.*

[17] James Gibson, Bo Xiao, Panayiotis G Georgiou, and Shrikanth Narayanan. 2013. An audio-visual approach to learning salient behaviors in couples' problem solving discussions. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW).* IEEE, 1–4.

[18] John Mordechai Gottman. 2005. *The mathematics of marriage: Dynamic nonlinear models.* MIT Press.

[19] John Mordechai Gottman. 2014. *What predicts divorce?: The relationship between marital processes and marital outcomes.* Psychology Press.

[20] John M Gottman and Robert W Levenson. 1985. A valid procedure for obtaining self-report of affect in marital interaction. *Journal of consulting and clinical psychology* 53, 2 (1985), 151.

[21] John M Gottman and Robert W Levenson. 1986. Assessing the role of emotion in marriage. *Behavioral Assessment* (1986).

[22] Shawn Hershey, Sourish Chaudhuri, Daniel PW Ellis, Jort F Gemmeke, Aren Jansen, R Channing Moore, Manoj Plakal, Devin Platt, Rif A Saurous, Bryan Seybold, et al. 2017. CNN architectures for large-scale audio classification. In *2017 ieee international conference on acoustics, speech and signal processing (icassp).* IEEE, 131–135.

[23] Richard E Heyman. 2001. Observation of couple conflicts: Clinical assessment applications, stubborn truths, and shaky foundations. *Psychological assessment* 13, 1 (2001), 5.

[24] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[25] Daniel Kahneman. 2000. Evaluation by moments: Past and future. *Choices, values, and frames* (2000), 693–708.

[26] Athanasios Katsamanis, James Gibson, Matthew P Black, and Shrikanth S Narayanan. 2011. Multiple instance learning for classification of human behavior observations. In *International Conference on Affective Computing and Intelligent Interaction.* Springer, 145–154.

[27] Patricia K Kerig and Donald H Baucom. 2004. *Couple observational coding systems.* Taylor & Francis.

[28] Chi-Chun Lee, Athanasios Katsamanis, Matthew P Black, Brian R Baucom, Panayiotis G Georgiou, and Shrikanth Narayanan. 2011. An analysis of PCA-based vocal entrainment measures in married couples' affective spoken interactions. In *Twelfth Annual Conference of the International Speech Communication Association.*

[29] Chi-Chun Lee, Athanasios Katsamanis, Matthew P Black, Brian R Baucom, Panayiotis G Georgiou, and Shrikanth S Narayanan. 2011. Affective state recognition in married couples' interactions using PCA-based vocal entrainment measures with multiple instance learning. In *International Conference on Affective Computing and Intelligent Interaction.* Springer, 31–41.

[30] Robert W Levenson, Laura L Carstensen, and John M Gottman. 1994. Influence of age and gender on affect, physiology, and their interrelations: A study of long-term marriages. *Journal of personality and social psychology* 67, 1 (1994), 56.

[31] Robert W Levenson and John M Gottman. 1983. Marital interaction: physiological linkage and affective exchange. *Journal of personality and social psychology* 45, 3 (1983), 587.

[32] Angeliki Metallinou, Chi-Chun Lee, Carlos Busso, Sharon Carnicke, Shrikanth Narayanan, et al. 2010. The USC CreativeIT database: A multimodal database of theatrical improvisation. *Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality* (2010), 55.

[33] Angeliki Metallinou and Shrikanth Narayanan. 2013. Annotation and processing of continuous emotional attributes: Challenges and opportunities. In *2013 10th IEEE international conference and workshops on automatic face and gesture recognition (FG).* IEEE, 1–8.

[34] Philipp M Müller, Sikandar Amin, Prateek Verma, Mykhaylo Andriluka, and Andreas Bulling. 2015. Emotion recognition from embedded bodily expressions and speech during dyadic interactions. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII).* IEEE, 663–669.

[35] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. 2015. Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction.* 443–449.

[36] Sally Olderbak, Andrea Hildebrandt, Thomas Pinkpank, Werner Sommer, and Oliver Wilhelm. 2014. Psychometric challenges and proposed solutions when scoring facial emotion expression codes. *Behavior Research Methods* 46, 4 (2014), 992–1006.

[37] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.

[38] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* 37 (2017), 98–125.

[39] Anna Marie Ruef and Robert W Levenson. 2007. Continuous measurement of emotion. *Handbook of emotion elicitation and assessment* (2007), 286–297.

[40] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[41] James A Russell, Anna Weiss, and Gerald A Mendelsohn. 1989. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology* 57, 3 (1989), 493.

[42] Sourav Sahoo, Puneet Kumar, Balasubramanian Raman, and Partha Pratim Roy. 2019. A Segment Level Approach to Speech Emotion Recognition Using Transfer Learning. In *Asian Conference on Pattern Recognition.* Springer, 435–448.

[43] Laura Sels, Jed Cabrieto, Emily Butler, Harry Reis, Eva Ceulemans, and Peter Kuppens. 2019. The occurrence and correlates of emotional interdependence in romantic relationships. *Journal of personality and social psychology* (2019).

[44] Laura Sels, Eva Ceulemans, and Peter Kuppens. 2019. All's well that ends well? A test of the peak-end rule in couples' conflict discussions. *European Journal of Social Psychology* 49, 4 (2019), 794–806.

[45] Laura Sels, Yan Ruan, Peter Kuppens, Eva Ceulemans, and Harry Reis. 2020. Actual and perceived emotional similarity in couples' daily lives. *Social Psychological and Personality Science* 11, 2 (2020), 266–275.

[46] Tessa V West and David A Kenny. 2011. The truth and bias model of judgment. *Psychological review* 118, 2 (2011), 357.