HMM-based Detection of Head Nods to Evaluate Conversational Engagement from Head Motion Data

Saygin Artiran*, Leanne Chukoskie†, Ara Jung†, Ian Miller†, Pamela Cosman*
*Department of Electrical and Computer Engineering, †Qualcomm Institute
University of California, San Diego
{sartiran, lchukoskie, a4jung, pcosman}@ucsd.edu, ian@iantmiller.com

***Abstract*—Head gestures such as head nodding and shaking play a prominent role in conversation, indicating active listening and interest in a conversation. We aim to create a tool to assess these physical conversational engagement cues in the context of a mock job interview. We propose a hidden Markov model-based architecture to locate and classify head nods and shakes from head motion data in an online fashion. Based on the number, velocity, and duration of the detected head gestures, we evaluate the conversational engagement level using a linear regression model. For the interview segments, high agreement was reached between model scores and scores from human raters. We consider this system as a path toward augmented reality and virtual reality-based training that can broaden participation in careers with competitive hiring scenarios.**

***Index Terms*—Conversational engagement, job interview training, head motion, hidden Markov models.**

I. Introduction

Technologies are needed that support assessment and learning of social communication skills. Cues that indicate active listening include physical backchannels such as head nodding and shaking, facial expressions, and oral backchannels, such as saying "yeah" and "mm-hmm". Conversational engagement behavior is atypical in individuals with autism spectrum disorder (ASD). Studies report that 1 in 45 individuals is diagnosed with ASD and only 15% are employed [1]. While many adults with ASD have the cognitive skill to contribute substantially to the workforce, social communication deficits prevent many from obtaining employment.

Here we report on a system that uses an augmented reality (AR) headset in the context of a social interaction (an interview) to score conversational engagement from head motion data. This approach would translate well to a virtual reality (VR) headset in which the subject interacts with a virtual interviewer. Virtual-based instruction has multiple advantages from a training perspective as scenarios are repeatable, controllable and individualizable [2]. This is evidenced in social skills research, including teaching interpersonal communication and interview skills. In a review, Bonaccio and colleagues [3] investigate nonverbal behavior in the workplace and identify many "codes" of nonverbal behavior, from kinesics (communication through body movement), to haptics (communication through touch), vocalics (communication through voice), and proxemics (communication through physical space). In terms of interview specific skills, Arvey and Campion [4] found that nonverbal behaviors like eye gaze and body movement significantly influence how an interviewer perceives an interviewee's performance. These perceptions, then, bias the selection of new employees.

This work was supported by the National Science Foundation under grant DUE-1928604.

Culture affects the meaning and execution of head gestures. For example, the "head bobble" is common in South Asia, notably in India [5]; this side-to-side head tilting is similar to a head shake but is used in place of a nod to show compliance. So, the models developed here for training conversational engagement cues have obvious cultural limitations but are also readily extensible through culturally-appropriate modules.

In this work, subjects wearing an AR headset respond to a mock interviewer who follows a script. The conversational engagement of the subjects is scored by three raters. The head motion data is input to hidden Markov models (HMMs) to detect head nods and shakes. The number, velocity and duration of head gestures are used in a linear regression step to predict the rater scores. The contributions of this paper are that we build (1) an HMM structure for detection of head nodding and shaking from inertial measurement unit (IMU) data in the context of subjects who move freely during a conversation, and (2) an evaluation model that scores physical backchannels during a mock interview, and has high agreement with raters.

II. Related Work

As we aim to develop a system to assist individuals in practicing and developing their conversational interaction skills, we use IMU data from a head-mounted display (HMD) so that the system could function with a virtual interviewer in a VR setup or with an in-person coach in an AR setup. While most past research on head gesture detection involves systems that track a subject's face via an external camera, some past work uses depth sensors or HMDs to directly exploit 3D head motion data. Yi *et al.* [6] built a similarity-based model using gyroscope and accelerometer data from smart glasses. Dynamic Time Warping (DTW) was used to recognize certain head movements (e.g., nod/shake, tilting, circling). Aiming to replace in-air hand commands of the Microsoft HoloLens glasses with head gestures, Yan *et al.* [7] used DTW to produce similarity scores between collected data sequences and a target head gesture set. In [8], Google Cardboard VR glasses were used to build a DTW-based system that detects certain head movements based on IMU data. Zhao and Allison

[9] implemented a cascaded-HMM architecture to recognize simple head movements from angular velocity data collected in real time using the Oculus Rift DK2. The structure consisted of nine discrete left-right HMMs. Severin [10] placed an inertial sensor on an audio headset and tracked the IMU data (roll, pitch, yaw) to detect 8 custom head movements.

Virtual training scenarios traditionally used 2D screens, e.g., Baur *et al.* [11] and Schouten *et al.* [12] collected nonverbal social cues, including hand movements, head direction, and facial expressions. The findings in these studies comport with previous research like [3], [4] which significantly weights the influence of nonverbal behaviors on interpersonal communication. The growing body of research on virtual based training systems supports the superiority of HMD systems (encompassing both VR and AR) over 2D methods. The authors of [13] used Google Glass to track autistic individuals' head movements to provide real-time feedback on gaze direction and speaking volume during an interview. The guidance given recognized that direct gaze might be uncomfortable, but some situations demand it. In [14], researchers used the Microsoft Kinect and physiological sensors, and measured college students' facial expressions and anxiety levels as measures of interview performance. Both studies indicated that the AR system can help certain groups cope with interview anxiety.

Aimed at young professionals, the system in [15] integrated an HTC Vive Pro VR headset with chat-bots, facial recognition software, physiological monitoring and emotion recognition from text responses to evaluate a user's interview performance. In another study, Taupiac *et al.* [16] created role-playing scenarios for managers and sales representatives, again using the HTC Vive Pro. These studies do not use head motion data.

One population that faces challenges in these areas are those on the autism spectrum. Didehbani *et al.* [17] taught social skills to autistic children using a VR HMD and the online platform *SecondLife*. Researchers measured emotion recognition, social attribution, and attention via tests before and after VR-based training. It is interesting to note the researchers' use of traditional assessment tools for a nontraditional VR-based training. Although such tools have validated construct validity, using static attention assessments to evaluate the success of this type of training could easily miss improved dynamic attention allocation. An analysis system like the one proposed in Section III could provide more objective measures of a user's attentiveness and dynamic attention allocation.

III. Head Gesture Classification using HMMs

Here we describe the data collection and model building for head gesture classification. Section IV describes the process for scoring conversational engagement.

A. Data Collection

The Magic Leap One AR headset produces (x, y, z) head position data at 60 samples/s. Ten adult subjects (7 men, 3 women) wearing the headset were asked to respond to 72 questions with head nods or shakes accompanied by yes or no oral answers. To elicit a range of behaviors, the questions included short items ("Do you live on campus?"), longer ones ("If my cousin was half of my age when I was 8, he will be 36 when I am 40, is that true?") and ones that might prompt a smile or some puzzlement ("Russian astronauts successfully landed on the Sun in 2018, is that correct?"). Answer correctness was not a concern, and subjects were asked to guess if needed. The headset recorded video, audio and 3D head position data during each session.

The videos were manually annotated to locate the start and end of each relevant head gesture. The longest, shortest, and average head gesture sequences had 193 samples (3.22 s), 36 samples (600 ms), and 95 samples (1.58 s). For online validation, we separated the video sections corresponding to the last 8 gestures. Of the 720 labeled gesture sequences, 82 were removed in a data cleaning step because they corresponded to cases where (a) the subject verbally answered but forgot to execute a head motion, (b) the gesture resembled a "head bobble" in which the motion was clearly similar to a head shake but was associated with an affirmative answer, and (c) there was a large body movement, as in cases of distraction. After data cleaning, we were left with 560 training gesture sequences (302 nods, 258 shakes) and 78 validation gesture sequences. We also extracted 284 non-gesture sequences from the same data; consisting of slight rotations in various directions as well as the idle state, this set was used as the negative or "neither" class while training.

B. System Architecture

The system begins with an HMM structure (Fig. 1) containing three parallel HMMs; Ψ_1 and Ψ_{-1} are trained to detect head nods and shakes, and Ψ_0 aims to detect everything else. In the Magic Leap 3D data, x is the horizontal left-right axis, y is the vertical axis, and z is the horizontal forward-backward axis. The HMM inputs are x and y velocity values, that is the position changes in x and y between consecutive samples (z components were discarded because they were found in the validation step to provide no benefit). In Fig. 1, Ψ_1, Ψ_0 and Ψ_{-1} are all left-right HMMs. The system processes sequences of x and y velocity values in overlapping windows. The HMM structure involves a hyperparameter set consisting of $\{M_1$, N_1, M_0, N_0, M_{-1}, N_{-1}, w_{size}, $s_{size}\}$, where M_i and N_i are the numbers of observation symbols and hidden states in HMM i, and w_{size} and s_{size} are the window size and step size of the sliding window. The velocity values go through separate quantization in each HMM since the numbers of clusters (M_i) might be different. The quantization discretizes the raw velocities using the centroids obtained during training, and the minimum distance classifier. After obtaining a discrete observation sequence (O), the Forward algorithm calculates the related log-likelihoods ($l(O \mid \Psi_i)$). The class of the HMM that produces the highest log-likelihood which is also greater than the associated minimum training log-likelihood determines the head movement type in that window. If there is a tie between two log-likelihoods, or the maximum log-likelihood is not greater the related minimum training log-likelihood, the model classifies a window as "neither".

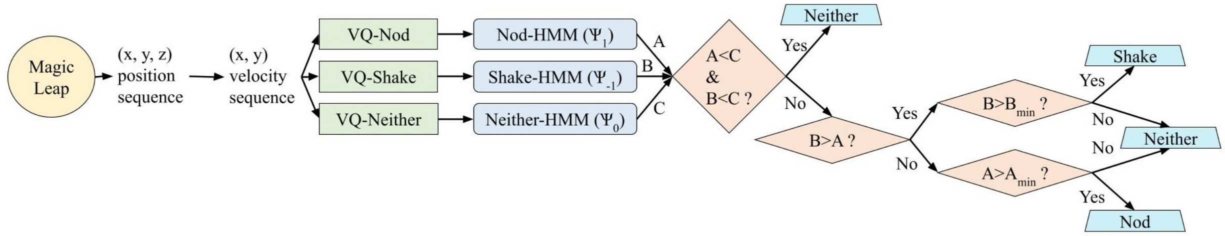

Fig. 1 HMM Structure: VQ stands for vector quantization. A, B, C are the log-likelihoods from the HMMs for the Nod, Shake, and Neither classes, and A_{min}, B_{min}, and C_{min} are the minimum log-likelihoods achieved for them during training.

Each window classification is associated with a confidence, defined as a ratio of log-likelihoods. If a window (W) is labeled as a head nod, the related classification confidence is:

$$C = \frac{l(W \mid \Psi_1)}{l(W \mid \Psi_1) + l(W \mid \Psi_0) + l(W \mid \Psi_{-1})}.$$

Here, $l(W \mid \Psi_1)$, $l(W \mid \Psi_0)$ and $l(W \mid \Psi_{-1})$ are the log-likelihoods for the head nod, neither, and head shake classes.

As a head gesture might get separated into several windows of which one or more might be misclassified, the complete sequence of class labels is filtered to correct errors. A run length (RL) filter with two parameters, entry (*T1*) and exit (*T2*), operates on two classes, head gesture and non-gesture. If the classification result is a head gesture for at least *T1* consecutive windows, those windows are considered as the start of a head gesture. A head gesture ends when the label is "neither" for at least *T2* consecutive windows. Variations on this type of RL filter have been used in applications involving temporal behavioral data, including gaze data for detecting looks to objects [18], and human activity recognition [19]. Each run length of windows considered to be a head gesture is labeled as a nod or shake based on the majority vote of the window labels; ties are resolved using the label of the highest confidence window. As an example, with *T1* = 1 and *T2* = 2, a sequence of labels $S = \{0, 0, 1, 1, 0, -1, 1, 0, -1, 0, 0\}$ would produce output $\hat{S} = \{0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0\}$.

1) Training: As in [9], the raw input vectors (velocity 2-tuples in our case) are vector quantized using the k-means algorithm with the number of clusters chosen to equal the number of velocity observation symbols. After obtaining the quantized observation sequences (O) from the training sequences, the Baum-Welch algorithm is used to train each HMM with the sequences from the associated class. The minimum training log-likelihoods for Ψ_1, Ψ_0 and Ψ_{-1} are -94.62, -105.32 and -89.79, while the maximum training log-likelihoods are -17.92, -15.58 and -13.56.

2) Validation: While the HMMs are trained with annotated gesture sequences, we perform online validation, meaning that a sliding window extracts and classifies data segments from a recording. The window shifts by s_{size}, and the parallel HMM structure processes the next window. Online validation allowed us to evaluate the system on many kinds of non-gesture head movements which might not have been encountered in manually annotated "neither" sequences. We evaluated values for N_i from 2 to 10, for M_i from 2 to 50, for w_{size} from 35 to 65, for s_{size} from $w_{size} - 30$ to $w_{size} - 10$, and for *T1* and *T2* from 1 to 10. For optimizing the hyperparameters, a possible objective function is the macro-averaged *F1* score (harmonic mean of the multi-class precision and multi-class recall). For scoring conversation engagement levels, head gesture duration may be relevant, for which the Intersection over Union (IoU) metric is useful. We optimized based on the sum of F1 and IoU, leading to the hyperparameter set $\hat{\Gamma} = \{\hat{M}_1 = 34, \hat{N}_1 = 7, \hat{M}_0 = 10, \hat{N}_0 = 7, \hat{M}_{-1} = 27, \hat{N}_{-1} = 7, \hat{w}_{size} = 38, \hat{s}_{size} = 28, \hat{T}1 = 1, \hat{T}2 = 2\}$. The validation results are in Table I. Here, it is not meaningful to report the number of "neither" gestures since it is arbitrary to say how many overlapping "neither" windows constitute a "neither" gesture. The precision, recall and *F1* score are 88.51%, 98.72% and 93.34%, respectively. Since the processing happens online using sliding windows, one true head gesture sequence can be falsely detected as multiple sequences. While calculating the evaluation metrics, for such mistakes we label only the longest sequence as a true positive, regarding the rest as false positives.

TABLE I: Classification Results for the Best Configuration

True \ Prediction	Head Nod	Head Shake	Neither
Head Nod	36	0	0
Head Shake	0	41	1
Neither	3	7	N/A

The IoU value is 64.80%, largely due to detection errors rather than duration errors. For the 78 true head gestures in our validation data, there are 10 false positives and 1 false negative. Excluding those from the IoU and examining the mismatches for true positives, we find the average absolute error in start (end) time is 11 (18) samples. These offsets are smaller than both the window size (w_{size}) and step size (s_{size}); such errors are inevitable with a sliding window approach as windows will not perfectly overlap with true sequence locations.

IV. Scoring Engagement Levels in Interviews

A. Data Collection and Ground Truth

We created a job interview script which takes about 10 minutes to complete. The mock interviewer is scripted, and the respondent is unscripted. The script includes questions about the characteristics and experience of the interviewee, and parts where the interviewer describes the lab and the projects on which the interviewee might work. During these descriptions, there are opportunities for the interviewee to nod in response, where the interviewer provides backchannel prompts such as "isn't that right?" or "you know what I mean?".

Fifteen adult subjects (9 men, 6 women) took part in the mock interviews. Each was interviewed twice and was asked to perform two different engagement levels. A subject enacting high engagement might show excitement, give many easily audible responses, and provide animated head nodding or shaking. An interviewee enacting low engagement may be passive and perform few head gestures. We divided the script into 6 parts, resulting in 180 data recordings (15 subjects $\times$ 2 times $\times$ 6 parts). One part provided 5 head gesture opportunities, three contained 6 opportunities, and the remaining two parts had 7 opportunities. These opportunities correspond to yes/no questions as well as backchannel prompts. Even without a specific question or prompt, the interviewee can choose to nod along during the descriptive portions, and can deliver their own short answers with more or less physical backchanneling.

During each interview, audio and video are recorded from an external camera. Interview segments are scored by three raters based on the video visuals, ignoring the answer content. Our scoring system is: (1) Head movements are absent or nearly absent. The subject appears either disengaged or impassive with low responsiveness. (2) Occasional head gestures. The subject may appear attentive but with lower than average responsiveness. (3) Typical or average engagement level for a two-way conversation with some head gestures. (4) Frequent head gestures, with above average engagement. (5) Lots of head gestures. The subject seems bouncy and enthusiastic. Raters could give intermediate values (1.5, 2.5, etc.) when the behavior seemed intermediate between two descriptions.

B. Scoring Model

The scoring model consists of the above HMM architecture and filtering, together with a feature extractor and a linear regression model that uses extracted head gesture features to predict an engagement score. To choose features and fit the model, we use cross-validation (CV), leaving out one subject at a time. As a preprocessing step in each CV fold, we use min-max normalization, normalizing both training and test data using the maximum and minimum values of each feature in the training set. We have 15 subjects with 12 interview segments each. Each interview segment is scored by three raters, so a segment has three labels. We fit the model to the data from 14 subjects which forms a training set of 504 interview segments, and test on the 12 unique segments from the subject left out.

We consider the following features: Total number of detected head gestures in a segment, Mean velocity (cm/sample) and Mean maximum velocity (cm/sample) and Duration (s) of the detected head gestures, Average head movement velocity (cm/sample) in the parts that are not classified as a head gesture. For each test segment, we compare the model output and the ground truth (average rater score) using Mean Absolute Error (MAE). We fit the regression model based on different feature subsets using subject-wise CV. The most descriptive feature on its own is the number of detected head gestures

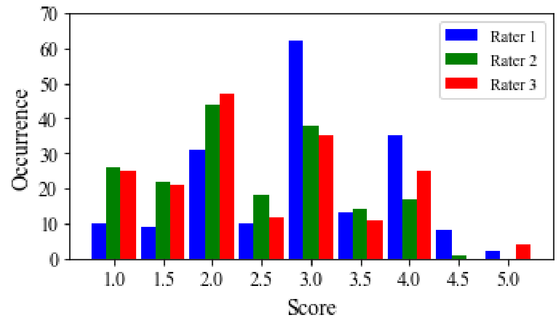

Fig. 2 Distribution of scores given by three raters

(MAE = 0.445, standard deviation = 0.344). The best feature pair is the number and average duration of detected head gestures (MAE = 0.373, s.d. = 0.292), and adding the average velocity maintains the MAE at 0.373, and slightly reduces the s.d. to 0.287. We do not use the other candidate features as they produce slightly worse MAE. The regression model is:

$$score = \alpha \cdot gst_{count} + \beta \cdot gst_{avg.vel.} + \omega \cdot gst_{avg.dur} + \eta.$$

V. Results

Fig. 2 depicts the distribution of rater scores. Raters 1 and 3 have experience working with individuals with ASD, however raters 2 and 3 produce similar scores, which tend to be lower than those of rater 1. The average differences between scores given by raters 1 and 2, 1 and 3, and 2 and 3 are 0.62, 0.62 and 0.37, respectively (0.54 on average). With subject-wise CV to fit the model, we get 15 different models, each with a different $\{\alpha, \beta, \omega, \eta\}$ set; the average values across the CV folds are $\{\hat{\alpha}$ = 2.35, $\hat{\beta}$ = 0.87, $\hat{\omega}$ = 1.16, $\hat{\eta}$ = 1.27$\}$ with standard deviations of 0.054, 0.064, 0.080 and 0.025.

Fig. 3 presents a scatter plot of model prediction values vs. average rater scores for each interview segment. Purple lines delimit the area where the absolute prediction error is below 0.5 (73.9% of points) while yellow lines correspond to error less than 1 (97.2% of points). The mean absolute prediction error is 0.37 (s.d. = 0.29). The maximum over and under estimations are 1.58 and 1.14.

Fig. 4 shows the cumulative distributions of prediction errors for different binned average rater scores (e.g., average

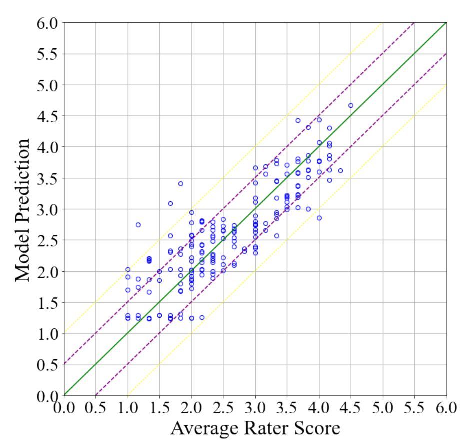

Fig. 3 Average rater score vs. model prediction for 180 interview segments

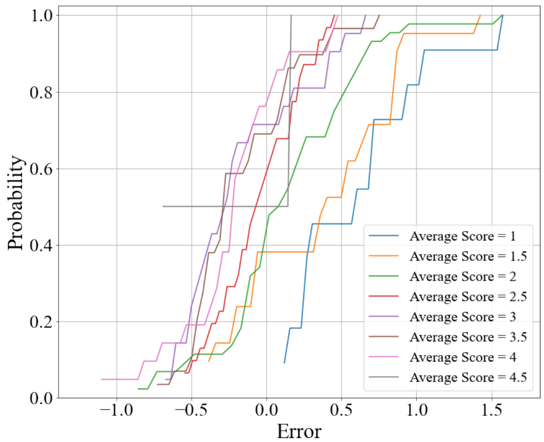

Fig. 4 Cumulative distribution curves of errors for each engagement level

scores between 4.25 and 4.75 go in the bin for 4.5). The plot does not contain a line for level 5, as the average score never exceeded 4.75. The model produces the largest errors for very low engagement examples; this is potentially not a significant drawback since whether a user receives a low or very low score, they would be aware of the need to display head gestures more frequently and enthusiastically. The model has higher accuracy in the mid to high range, allowing users to calibrate to a range of typical to high conversational engagement levels.

VI. Conclusion

Neurodivergent employees bring unique perspectives to teams, yet hiring these employees remains a challenge due to mismatched expectations about what constitutes engaged social conversation appropriate for an interview setting [20]. While multiple efforts are underway to change hiring practices to be more inclusive [20], we need better tools to help neurodivergent individuals practice normative social conversation skills. To prepare job applicants for job interviews, we built an AR headset-based application which tracks head position, detects head nods and shakes, and scores conversational engagement in terms of physical backchannels. The number of detected head gestures in a short interview segment, along with gesture velocity and duration, provide enough information to build a simple model that scores similarly to human evaluators.

While head gesture detection is widely studied, few studies use head motion data, and most that do [6]–[10] are in a context where the user aims to execute specific control gestures (a type of human computer interface) rather than moving freely in a naturalistic conversation. The context of a mock job interview naturally affects interviewees' head movements, however arbitrary head movements (e.g., looking away from the interviewer) were present in our data. The two main contributions of this paper are therefore performing head gesture detection with high accuracy in an online fashion while allowing subjects to move freely, and showing that extracted head gesture information in a simple model allows prediction of conversational engagement that has high agreement with human raters. For future work, an automatic rating system integrated with a VR platform would allow for solo practice.

References

[1] R. Cameto, C. Marder, M. Wagner, and D. Cardoso, "Youth employment," *NLTS2 Data Brief*, vol. 2, no. 2, pp. 1–6, 2003.

[2] X. Pan and A. F. d. C. Hamilton, "Why and how to use virtual reality to study human social interaction: the challenges of exploring a new research landscape," *British Journal of Psychology*, vol. 109, no. 3, pp. 395–417, 2018.

[3] S. Bonaccio, J. O'Reilly, S. L. O'Sullivan, and F. Chiocchio, "Nonverbal behavior and communication in the workplace: a review and an agenda for research," *J. of Management*, vol. 42, no. 5, pp. 1044–1074, 2016.

[4] R. D. Arvey and J. E. Campion, "The employment interview: a summary and review of recent research," *Personnel Psychology*, vol. 35, no. 2, pp. 281–322, 1982.

[5] P. Pillai, "Cultural directions and origins of everyday decisions," *Integrative Psychological and Behavioral Science*, vol. 46, no. 2, pp. 235–242, 2012.

[6] S. Yi, Z. Qin, E. Novak, Y. Yin, and Q. Li, "GlassGesture: exploring head gesture interface of smart glasses," in *IEEE International Conference on Computer Communications*. IEEE, 2016, pp. 1–9.

[7] Y. Yan, C. Yu, X. Yi, and Y. Shi, "HeadGesture: hands-free input approach leveraging head movements for HMD devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 4, pp. 1–23, 2018.

[8] T. Hachaj and M. R. Ogiela, "Head motion classification for single-accelerometer virtual reality hardware," in *International Conference on Frontiers of Signal Processing*. IEEE, 2019, pp. 45–49.

[9] J. Zhao and R. S. Allison, "Real-time head gesture recognition on head-mounted displays using cascaded hidden Markov models," in *IEEE Intl. Conf. on Systems, Man, and Cybernetics*. IEEE, 2017, pp. 2361–2366.

[10] I.-C. Severin, "Time series feature extraction for head gesture recognition: considerations toward HCI applications," in *Intl. Conf. on System Theory, Control and Computing*. IEEE, 2020, pp. 232–237.

[11] T. Baur, I. Damian, P. Gebhard, K. Porayska-Pomsta, and E. André, "A job interview simulation: social cue-based interaction with a virtual character," in *International Conference on Social Computing*. IEEE, 2013, pp. 220–227.

[12] D. G. Schouten, F. Venneker, T. Bosse, M. A. Neerincx, and A. H. Cremers, "A digital coach that provides affective and social learning support to low-literate learners," *IEEE Transactions on Learning Technologies*, vol. 11, no. 1, pp. 67–80, 2017.

[13] Q. Xu, S.-c. S. Cheung, and N. Soares, "LittleHelper: an augmented reality glass application to assist individuals with autism in job interview," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, 2015, pp. 1276–1279.

[14] X. Jin, Y. Bian, W. Geng, Y. Chen, K. Chu, H. Hu, J. Liu, Y. Shi, and C. Yang, "Developing an agent-based virtual interview training system for college students with high shyness level," in *IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE, 2019, pp. 998–999.

[15] I. Stanica, M.-I. Dascalu, C. N. Bodea, and A. D. B. Moldoveanu, "VR job interview simulator: where virtual reality meets artificial intelligence for education," in *Zooming Innovation in Consumer Technologies Conference*. IEEE, 2018, pp. 9–12.

[16] J.-D. Taupiac, N. Rodriguez, O. Strauss, and P. Beney, "Social skills training tool in virtual reality, intended for managers and sales representatives," in *IEEE Conference on Virtual Reality and 3D User Interfaces*. IEEE, 2019, pp. 1183–1184.

[17] N. Didehbani, T. Allen, M. Kandalaft, D. Krawczyk, and S. Chapman, "Virtual reality social cognition training for children with high functioning autism," *Computers in Human Behavior*, vol. 62, pp. 703–711, 2016.

[18] L. Chukoskie, S. Guo, E. Ho, Y. Zheng, Q. Chen, V. Meng, J. Cao, N. Devgan, S. Wu, and P. C. Cosman, "Quantifying gaze behavior during real-world interactions using automated object, face, and fixation detection," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, pp. 1143–1152, 2018.

[19] G. Bhat, R. Deb, V. V. Chaurasia, H. Shill, and U. Y. Ogras, "Online human activity recognition using low-power wearable devices," in *IEEE/ACM Intl. Conf. on Computer-Aided Design*, 2018, pp. 1–8.

[20] A. Krzeminska, R. D. Austin, S. M. Bruyère, and D. Hedley, "The advantages and challenges of neurodiversity employment in organizations," *J. of Management & Organization*, vol. 25, no. 4, pp. 453–463, 2019.