

Head Movement Patterns during Face-to-Face Conversations Vary with Age

Denisa Qori McDonald
Children's Hospital of Philadelphia
Philadelphia, PA, USA
mcdonalddq@chop.edu

Casey J. Zampella
Children's Hospital of Philadelphia
Philadelphia, PA, USA
zampellac@chop.edu

Evangelos Sariyanidi
Children's Hospital of Philadelphia
Philadelphia, PA, USA
sariyanide@chop.edu

Aashvi Manakiwala
University of Pennsylvania
Philadelphia, PA, USA
Children's Hospital of Philadelphia
Philadelphia, PA, USA
aashvi@seas.upenn.edu

Ellis DeJardin
Children's Hospital of Philadelphia
Philadelphia, PA, USA
dejardine@chop.edu

John D. Herrington
Children's Hospital of Philadelphia
Philadelphia, PA, USA
University of Pennsylvania
Philadelphia, PA, USA
herringtonj@chop.edu

Robert T. Schultz
Children's Hospital of Philadelphia
Philadelphia, PA, USA
University of Pennsylvania
Philadelphia, PA, USA
schultzrt@chop.edu

Birkan Tunç
Children's Hospital of Philadelphia
Philadelphia, PA, USA
University of Pennsylvania
Philadelphia, PA, USA
tuncb@chop.edu

ABSTRACT

Advances in computational behavior analysis have the potential to increase our understanding of behavioral patterns and developmental trajectories in neurotypical individuals, as well as in individuals with mental health conditions marked by motor, social, and emotional difficulties. This study focuses on investigating how head movement patterns during face-to-face conversations vary with age from childhood through adulthood. We rely on computer vision techniques due to their suitability for analysis of social behaviors in naturalistic settings, since video data capture can be unobtrusively embedded within conversations between two social partners. The methods in this work include unsupervised learning for movement pattern clustering, and supervised classification and regression as a function of age. The results demonstrate that 3-minute video recordings of head movements during conversations show patterns that distinguish between participants that are younger vs. older than 12 years with 78% accuracy. Additionally, we extract relevant patterns of head movement upon which the age distinction was determined by our models.

CCS CONCEPTS

• **Applied computing** → **Psychology**; • **Computing methodologies** → **Computer vision**; **Machine learning**; *Cluster analysis*;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICMI '22 Companion, November 7–11, 2022, Bengaluru, India

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9389-8/22/11...\$15.00

<https://doi.org/10.1145/3536220.3563366>

Supervised learning by classification; Supervised learning by regression; Cross-validation.

KEYWORDS

video analysis, bag-of-words approach, head movement patterns, dyadic features, monadic features, behavioral analysis, conversation analysis, non-verbal communication

ACM Reference Format:

Denisa Qori McDonald, Casey J. Zampella, Evangelos Sariyanidi, Aashvi Manakiwala, Ellis DeJardin, John D. Herrington, Robert T. Schultz, and Birkan Tunç. 2022. Head Movement Patterns during Face-to-Face Conversations Vary with Age. In *INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION (ICMI '22 Companion)*, November 7–11, 2022, Bengaluru, India. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3536220.3563366>

1 INTRODUCTION

Nonverbal communication, carried out through a variety of modalities, including eye gaze, facial expression, gesture, posture, orientation, and other head and body movements [16], is a rich and complex domain of functioning in humans. Nonverbal cues are used alone, in combination with one another, and alongside speech cues throughout development in order to successfully communicate. The in-depth analysis of nonverbal cues holds important implications for understanding human behavior. Such analysis can characterize how we relate to others [3, 34] and how the ways we express ourselves vary as a function of age, sex, context, and culture [20, 39]. In-depth study of nonverbal communication also has the potential to improve understanding of mental health conditions characterized in part by nonverbal behavior differences. Nonverbal communication is a truly trans-diagnostic domain of

functioning with differences observed in several mental health conditions, including autism, depression, anxiety, schizophrenia, and ADHD [7, 9, 11, 13, 14, 21, 33].

Successful social communication is achieved through the execution and coordination of multiple nonverbal modalities (e.g., facial expressions, eye gaze, head movements). The current study focuses on the analysis of one critical modality: head movement. Head movements serve several important functions during social interaction, including conveying meaning, signaling engagement and turn-taking, and providing structure [6, 30]. Studying head movements may improve understanding of the phenomenon of interpersonal coordination [15] – the tendency for social partners to match the form and timing of one another’s verbal and nonverbal cues – which has been found to be critical for positive social interaction, and reduced in certain mental health conditions [8, 26, 41]. Unfortunately, the characteristics of head movements during social interactions are understudied relative to other nonverbal communication domains, particularly in age groups outside of infancy and toddlerhood.

Traditionally, the conversational dynamics of nonverbal behavior have been studied through manual annotations (e.g., moments of head nodding) by trained observers. Manual annotations [4, 12], however, can potentially lack scalability and reliability, or require a considerable time investment. Recent developments in computer vision and human–computer interaction have enabled more precise and automated analysis of several behavioral modalities [10, 22] using a variety of computational models. As a result of these advances, it is now possible to obtain a rich characterization of nonverbal behaviors as they unfold over time, including nuanced differences in how different cues are used across people.

In this study, we used a computational approach to investigate patterns of head movements used by neurotypical school-aged children, adolescents, and adults during face-to-face conversations with an unfamiliar adult, as well as how those movement patterns varied with age. Our goal is to enhance understanding of normative patterns of head movements over typical development, in order to establish developmentally-sensitive normative models, which in turn are critical for studying atypicality (i.e., deviation from those norms) and individual differences [40]. Historically, a lack of scalable tools for measuring fine-grained characteristics of head movements during natural social interactions has hindered understanding of how use of this essential component of social communication develops over the lifespan. In response to this knowledge gap, the aim of the current study was two-fold: 1) To determine if we can reliably capture and characterize head movements during natural conversations between interacting partners with unobtrusive technological methods (e.g., without the use of wearables or distracting data collection setups), and 2) To begin to map the extent to which age affects head movement patterns.

To minimize interference with the natural flow of conversation, data was collected through a specialized device developed in-house (Section 3.2), comprised of two cameras, recording in synchrony, with one facing the participant and the other facing the unfamiliar adult interaction partner (i.e., "confederate"). Analysis of head movement patterns in this study relies on unsupervised learning, as well as classification and regression relating to participant age. For classification, we defined two age groups: ages 5–12 years and ages

12–50 years. This decision was informed by previous work [33] indicating that these age groups accurately differentiated individuals with respect to other aspects of social communication during similar interactions, as well as limited evidence that children’s nonverbal communication behavior becomes more similar to adults during preadolescence [24].

There are three unique contributions of the current work. First, we studied head movements at two levels of analysis: *monadic* (i.e., relating to participant alone) and *dyadic* (i.e., relating to both participant and confederate). These two levels of analysis allowed us to better capture the broader social context within which the head movements occurred. Second, unlike most previous work (Section 2.3) that characterized head movements simply in terms of overall kinematic features (e.g., speed, spatial range), we modeled the way head movements unfold during conversations (i.e., patterns of movements). Third, we quantified the extent to which head movements are shaped by development, by measuring how accurately the age of a person could be predicted based on the ways in which they used head movements during social interactions.

2 RELATED WORK

2.1 Significance of Head Movement for Social Communication

Head movement is a fundamental aspect of nonverbal communication. Often, communicative head movement is thought of in terms of discrete head gestures, such as head nods and shakes, to communicate “yes” and “no”, respectively. However, human beings use their heads to communicate in a host of additional ways, and commonly pair head movement with other forms of communication for emphasis or clarity. For example, McClave [30] put forth a taxonomy of head movements based on micro-analysis of videotaped conversations among adults, detailing specific movements associated with conveying certain linguistic meanings, referencing space, and controlling interpersonal interaction (e.g., turn-taking, backchanneling). This work suggests that discrete patterns of head movement carry important signal within a conversation, highlighting the significance of measuring head movement both precisely, and within natural interactive paradigms.

The use of head movement may also provide information about mental and emotional characteristics of interaction partners. For example, head movements can vary based on personal attributes of the speaker or listener (e.g., sex [1]). Furthermore, head movements have been associated with emotional and cognitive states of an individual (e.g., agreement, interest [23]), emotional expression [27], emotional interpersonal dynamics (e.g., conflict between interaction partners [17], as well as emotional responsiveness between infants and mothers [18]). Despite this evidence that head movement is a critical social communication tool, there is a paucity of findings in the psychology literature on head movement relative to other nonverbal domains such as facial expression and gestures [18].

2.2 Head Movement Across Development

The development and function of head movement is fairly well-understood in infancy. Most infants develop increasing head control



Figure 1: Experimental setup and data collection hardware. In (a), the participant and confederate are having a conversation with the “BioSensor” camera (a slightly older model shown here) placed between them. In (b), the current model of “BioSensor” used for data collection is shown.

gradually over the first six months of life, and engage head movements to track and obtain objects in their environments. Throughout infancy and toddlerhood, head movements are also fundamental to important social developmental skills, like social orienting, and social referencing. There is computationally-derived evidence to suggest that coordination among infants’ and their mothers’ head movements facilitates emotional communication [18], and that young children’s head movements play an increasing role in their social interactions as they develop from 6 to 24 months of age [37].

The role and characteristics of head movements are less well-studied in older children, adolescents, and adults. Murray and colleagues [32] studied non-social head movements in children ages 4-15 years in response to visual stimuli during a specific oculomotor task. They found that both frequency and variability of head movement decreased linearly as a function of age, and concluded that children engage the head less during gaze shifts as they age. However, it is difficult to know how or if these findings might relate to the social use of head movement.

2.3 Computational Analysis of Head Movement

Recent advances in computer vision and machine learning promise to rapidly advance research on human behavior, by introducing reliable and granular measurement tools within a new paradigm: computational behavior analysis. Such tools capture and quantify all observable human behavior with extraordinary precision. Over the past several years, a growing body of work has demonstrated the promise of computational behavior analysis for enabling detailed analysis of nonverbal communication during screen-based tasks (e.g., participant watching videos or looking at images) [29] and, less commonly, live social interactions [15].

Early prominent works using computational approaches to measure human head movements, and nonverbal cues in general, come chiefly from the human-robot and human-computer interaction

domain [38, 42]. Detection of head pose and its temporal variation (*i.e.*, movement of head) has often been used to improve robots’ communicative behavior, by informing robots’ own head movements to make them more natural [31]. When robots are able to mimic human’s nonverbal communication patterns, they are perceived as more expressive and lifelike, which enhances how people interact with them [19].

Over the past several years, interest has grown in the study of head movement as a lens for understanding mental health and neurodevelopmental conditions, in addition to neurotypical social behavior patterns. Autism has been one particularly ripe use-case for computer vision-based head movement analysis, given that it is characterized by differences in both nonverbal social communication [2] and motor skills [43, 44]. Indeed, autistic children have been shown to exhibit differences in head movement dynamics relative to their neurotypical peers. For example, while watching brief social and non-social movies, toddlers who go on to receive an autism diagnosis demonstrate significantly higher rate, acceleration, and complexity in their head movements compared to toddlers developing typically [25]. In a slightly older sample (ages 2-6 years), autistic and neurotypical children were found to differ in terms of lateral, but not vertical, head displacement and velocity, and only while watching social stimuli [29]. Zhao and colleagues [45, 46] extended the study of head movement dynamics in autism to both older children (6-13 years) as well as a dyadic social context (*i.e.*, a conversation with an adult). They found that head movements were greater and more stereotyped in autism.

Beyond just having implications for autism, these results underscore the potential for head movement analysis to advance understanding of a range of clinical presentations. They also speak to a need to examine head movements across multiple stages of development, to truly parse what is typical and atypical for any given age group. A deeper examination of how head movement

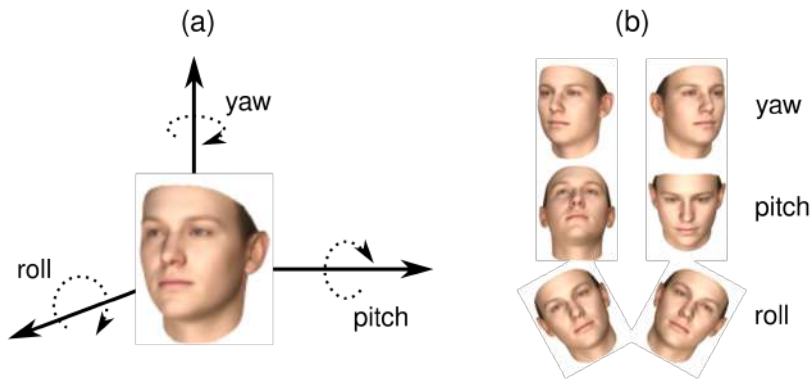


Figure 2: Illustration of the three angles used in the study to quantify head movement.

dynamics unfold over the course of typical development (*i.e.*, from childhood through adulthood), as well as their various social functions, will provide an important foundation for future studies in human behavior.

3 METHODS

3.1 Participants

Data collection was performed at the Center for Autism Research (CAR) at Children’s Hospital of Philadelphia (CHOP). Data collection and use was approved by the institutional review board (IRB) at CHOP. 79 individuals with no known mental health diagnosis (*i.e.*, neurotypical) are included in the current study. Absence of clinically significant mental health concerns was confirmed per evaluation by the clinical core at CAR, based on expert clinical judgement of DSM–V criteria [2]. Sample characteristics are summarized in Table 1.

Table 1: Sample characterization by age group

Age Group	n	Age (years) Min-Max	Sex F:M	Full Scale IQ Mean (Std)
≤ 12	39	5 – 12	18:21	111.6(13.8)
> 12	40	12 – 48	11:29	109.6(11.0)
Total	79	5 – 48	29:50	110.6(12.4)

3.2 Experimental Procedure

Participants underwent a battery of tasks to assess social communication competence. This battery included a modified version of the Contextual Assessment of Social Skills (CASS) [35]. The CASS is a semi-structured assessment of conversational ability designed to mimic real-life first-time encounters. Participants engaged in two 3-minute face-to-face conversations with two different confederates (unaware of the dependent variables of interest). CASS confederates included undergraduate students and research assistants (all native English speakers). Confederates were assigned to

participants based on availability. In the first conversation (interested condition), the confederate demonstrated social interest by engaging both verbally and non-verbally in the conversation. In the second conversation (bored condition), the confederate indicated boredom and disengagement verbally (*e.g.*, one-word answers, limited follow-up questions) and non-verbally (*e.g.*, neutral affect, limited eye-contact and gestures). The current analysis is based on the interested condition only.

Prior to each conversation, study staff provided the following prompt to the participants and confederates before leaving the room: “Thank you both so much for coming in today. Right now, you will have 3 minutes to talk and get to know each other, and then I will come back into the room.” In order to provide opportunities for participants to initiate and develop the conversation, confederates were trained to speak for no more than 50% of the time and to wait 10s to initiate the conversation. If conversational pauses occurred, confederates were trained to wait 5s before re-initiating the conversation. No additional instructions were provided to either speaker.

3.3 Data Collection

Continuous audio and video of the 3-minute CASS were recorded using a specialized “BioSensor”, built in-house (Figure 1), that was placed between the participant and confederate on a floor stand. This device has two HD video cameras pointing in opposite directions, as well as two microphones, to allow for simultaneous audio-video recording of the participant and the confederate as they sit facing each other. The minimal footprint and size of the device were intended to minimize the intrusiveness of the technology on the natural conversation.

3.4 Data Pre-Processing

An overview of the data processing and analysis pipeline is provided in Figure 3. For each video, the first and last 3 seconds were trimmed to remove any frames for which the research staff may have been in the room (*e.g.*, providing the instructions). This time limit was selected based on visual inspection of the videos. The majority of videos were recorded at a frame rate of 30 frames per second (fps), with some videos down-sampled from 60 fps for consistency.

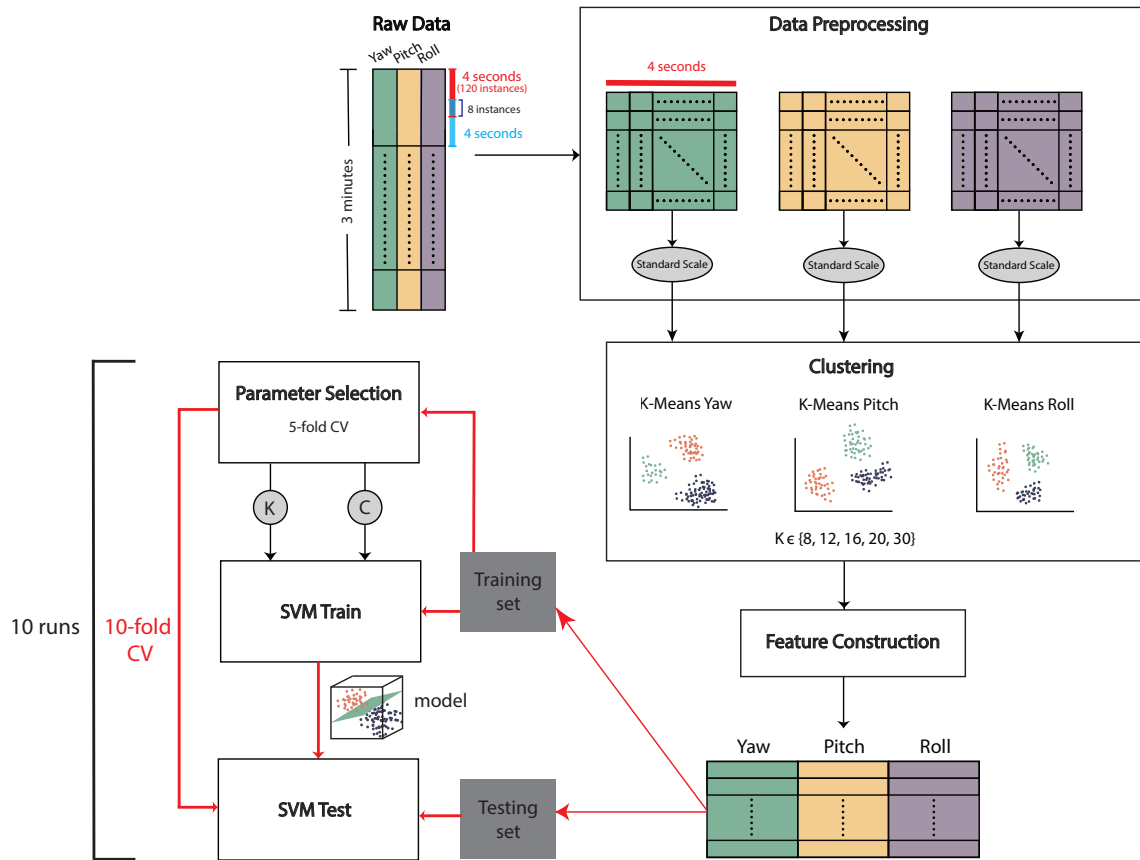


Figure 3: Overview of the data transformation processes and experiments. Each column of the raw data contains information about changes in each of the three basic head movement angles: yaw, pitch, and roll. Windowed data are produced, normalized to a standard scale (0 mean and 1 standard deviation), and analyzed for each angle independently. For each angle, windowed samples from all participants are clustered to identify common movement patterns. This step is repeated for multiple values of K in K-Means clustering. After the frequency-based feature construction (detailed in Figure 4), features from all angles are concatenated to produce the feature set for classification. Through 10-fold cross-validation, training data is also used to determine the best K for clustering and classifier parameters within a nested cross-validation framework. The trained model is then used for testing. Cross-validation is repeated 10 times with different random seeds for data shuffling (*i.e.*, 10 times 10-fold nested cross-validation).

Video data for all participants and confederates were processed using a state-of-the-art 3D face modeling algorithm [36] to extract features related to facial expressions and head pose. Only the latter were used in the current study. The output consisted of time-dependent signals for the three fundamental head movement dimensions: *yaw*, *pitch*, and *roll*, which are visualized in Figure 2. These signals were used in the rest of the analysis pipeline.

3.5 K-Means Clustering

We used K-Means clustering as a way to group head movement snapshots by similarity, with the goal of finding general patterns. We selected a K value that best fit the inherent patterns in the

data within a cross-validation framework (Section 3.7). We trained three separate K-Means models, for each of the *roll*, *pitch*, and *yaw* angles.

For each participant, the temporal signal of each angle was split into overlapping windows of 4 seconds (corresponding to 120 time instances given the frame rate of 30 fps), with an overlap size of 8 time instances (roughly 0.3 seconds). Each window was subsequently standardized to have 0 mean and 1 standard deviation. Windowed data from all participants (52,692 windowed instances) were combined together and used as input to train a K-Means model for the angle from which the window was generated.

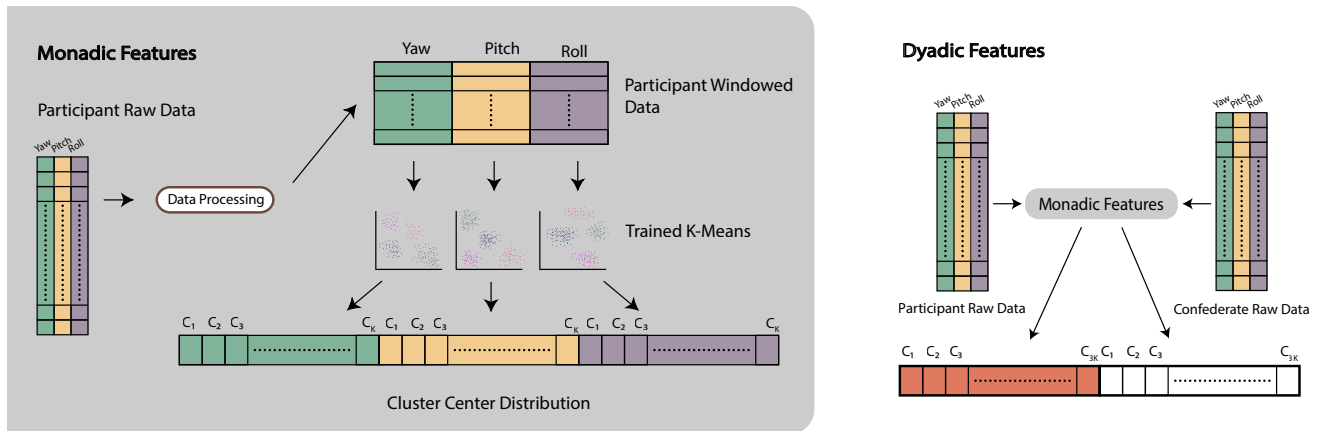


Figure 4: Illustration of the feature construction process. The *monadic* features are created by counting the number of windowed data that belong to each cluster. For each participant, windowed data is assigned to the closest cluster center, and the number of members in each cluster is counted. In the case of *dyadic* features, the monadic features are calculated both for the participant and the confederate, and concatenated to produce the final feature vector.

3.6 Feature Construction

The feature construction process, illustrated in Figure 4, consisted of counting the number of times each participant exhibited specific head movement patterns, represented by the trained K -Means models' cluster centers. We used two approaches for creating features: *monadic* and *dyadic*.

In the *monadic* case, only the participants' data was used. Data pre-processing happened as described in Section 3.4, producing windowed samples, and then clustering those samples. For each participant, the number of windowed samples belonging to each cluster was counted, and a K -dimensional vector of pattern counts was generated. Each participant was ultimately represented by a $3K$ -dimensional vector generated by concatenating pattern counts from the three different angles.

In the *dyadic* case, the *monadic* features were computed for both the participant and the confederate. The vectorized features of each were concatenated, resulting in a $6K$ -dimensional feature vector. The purpose of this strategy is to also capture the behavior of the conversation partner, as the *exchange* of nonverbal cues between two social partners is highly relevant for effective communication [5, 15].

3.7 Classification and Parameter Selection

The head movement features were used as the input to a binary SVM classifier with a linear kernel, to classify age (younger or older than 12 years). We used 10-fold cross-validation. For each fold, part of the data was put aside for testing, and the remaining was used for training and parameter selection. Parameter selection was done with an 'inner' 5-fold cross-validation (*i.e.*, nested cross-validation) using only training data. The parameters to select were the best K value for clustering, and the C parameter of the SVM classifier, which controls the amount of regularization in the model. The tested K values were {8, 12, 16, 20, 30}, and C values were {0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000}.

After the K and C values were selected, for the 'outer' cross-validation fold, we trained an SVM classifier with the specific C value, using the training dataset with features produced from the selected K value. The testing set with the features of that particular K value was then used to predict to which age group each testing-set participant belonged. At the end of the 10-fold cross-validation process, each participant's data had been used in the testing set once.

This 10-fold cross-validation was then repeated 10 times, with a different random seed used each time to shuffle the data, to produce statistically robust performance metrics. The K value selected the most times across all folds was stored to be used in the regression analysis (Section 3.8) and to identify relevant features (Section 3.9).

3.8 Regression

In addition to age classification, predicting a continuous value for age is also a relevant exploration. Similar to the classification explained above, we used a nested cross-validation, only fixing the K (using the value most used in classification) to facilitate easy comparisons between classification and regression. In addition to C , we also optimized the kernel choice (linear or rbf) within the 'inner' cross-validation.

3.9 Relevant Feature Extraction

Beyond determining simply whether head movement patterns distinguish between two age groups or predict continuous age, we sought to elucidate which head movement features contributed most strongly to classification and prediction accuracy. This analysis is potentially developmentally- and clinically-relevant, in that it may identify specific markers associated with typical head movement patterns at various ages, and inform future study of how head movements in non-neurotypical populations diverge from normative developmental patterns.

Table 2: Classification results for both *monadic* and *dyadic* features. The reported scores are means of the 10 experimental runs and balanced across the two classes.

	Accuracy	F1-Score	Precision	Recall	ROC (AUC)
<i>Monadic</i>	71.4%	71.4%	71.5%	71.4%	71.4%
<i>Dyadic</i>	78.6%	78.6%	78.7%	78.6%	78.6%

For this analysis, we used the K values that were selected as best-performing most frequently, both in the *monadic* and *dyadic* experiment sets, and re-ran classification using the entire dataset. Note that computing feature importance values within the nested cross-validation explained above would not be possible, since the K parameter and corresponding clusters (and therefore generated features and their meaning) change within each fold. Thus, after computing the performance metrics using the nested cross-validation and reporting them, we fixed all parameters including K , and ran a single classifier using the entire dataset. Subsequently, we extracted the feature coefficients computed by the algorithm for each input feature – they correspond to the importance of that feature in the classification process and are thus assumed to possess a high information content. Each of the extracted features corresponds to a specific head movement pattern of the participant in the *monadic* case, or either the participant or the confederate in the *dyadic* case. We then visualized the head movement patterns (*i.e.*, how a certain head angle changes over time) corresponding to those features.

4 RESULTS

4.1 Classification

These results relate to our investigation of whether it is possible to distinguish between participants belonging to younger (12 or below) or older (above 12) age groups. Data was analyzed through methods described in Sections 3.4 – 3.7, and the results are presented in Table 2. The accuracy, f1-score, precision, recall, and area under the curve (AUC) of the receiver operating characteristic (ROC) metrics were calculated for each of the 10 runs of cross-validation (Section 3.7). The values reported here are the means of those scores, balanced for each class; each metric was calculated while accounting for the number of instances per class.

All results were significantly better than chance, suggesting the existence of developmentally informative signals in the head movement patterns of participants. The *dyadic* approach achieved higher accuracy (78.6%, standard deviation of 2.1) than the *monadic* approach (71.4%, standard deviation of 3.5), highlighting the importance of studying social communication cues within a true social context (*i.e.*, also considering the behavior of the social partner).

4.2 Regression

To supplement the classification results and gain more insight into the relationship between head movement patterns and age, we also conducted a regression experiment, as detailed in Section 3.8. We used the best performing features in classification, which were those from the *dyadic* approach. Results yielded an r (Pearson correlation) between ground truth (chronological) age and predicted age of

0.42 ($p=0.0001$), indicating a moderately strong and statistically significant linear relationship.

4.3 Relevant Features

The K parameter was set to 16 (*monadic*) and 12 (*dyadic*) consistently in most folds of the nested cross-validation. Following the process described in Section 3.9, we extracted head movement patterns that were relevant in distinguishing between the two classes. Figure 5 gives the weights (*i.e.*, importance) of the most contributing features in the classification of participants into age groups. Note that the features quantify the number of times each head movement pattern (illustrated in Figure 6) was exhibited during the 3-minute conversation.

In the *monadic* case, the top 10 most informative features included movements in all three angles. Most features (8 out of 10) showed an increased frequency of occurrence in the younger group (12 years or younger). For example, the movement pattern c15 (the first feature in Figures 5 and 6), which corresponds to a sudden rotation of the head in the roll direction, occurred more in the younger group. Similarly, the only two yaw movements included in the list were more frequent in the younger group.

No movements in the yaw direction (*e.g.*, head shake) were included in the top 10 for the *dyadic* case. Notably, the most important *dyadic* features were mostly from confederates (8 out of 10). The younger and older groups showed higher frequencies in different patterns and directions.

5 DISCUSSION AND FUTURE WORK

We used computer vision and multivariate machine learning methodologies to study the head movements of interacting partners. The experiments and results introduced in this work demonstrate progress toward understanding the developmental characteristics of head movements during natural social interactions. As a first step, the current work provides metrics (*i.e.*, classification and regression accuracy) to quantify the extent to which head movements change with age.

Our results suggest that it is possible to distinguish between individuals younger versus older than 12 years based on their head movement patterns during a 3-minute, casual, face-to-face conversation. Within the age range of our sample (5.5-48 years), head movements could also be used to accurately predict age continuously. Furthermore, our experiments show that the *dyadic* approach achieved higher accuracy than the *monadic* approach, suggesting the possibility that the age of a person not only affects the head movement patterns that they themselves use during conversations, but also those used by conversation partners. These findings highlight the importance of studying social communication—as the name implies—within natural social interaction contexts, because

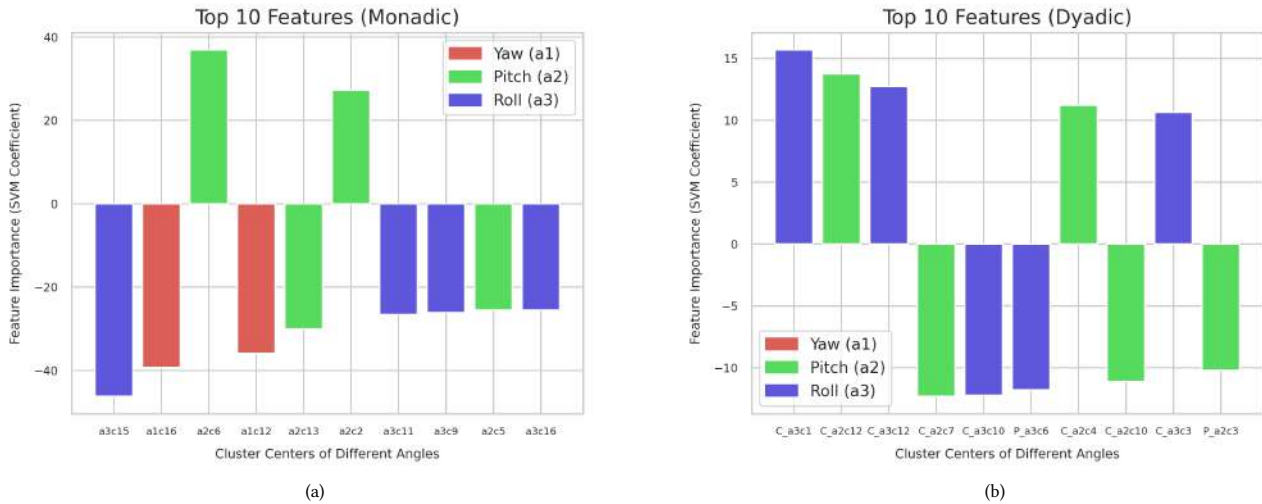


Figure 5: The top 10 most relevant features extracted during classification have been plotted for each feature construction scenario: *monadic* in (a) and *dyadic* in (b). For each feature, the y-axis represents the weight coefficients assigned by the SVM classifier. In the *monadic* case, each feature contains information about the angle (a) and cluster center within that angle (c). The *dyadic* feature plot contains the extra information of whether that feature belonged to the confederate (C) or the participant (P).

characteristics of a person (*i.e.*, age) influence both their own behavior and how other people behave in response. Computational behavior analysis, with its superior precision and scalability compared to manual annotations, facilitates such naturalistic studies.

While our results suggest that it is possible to predict age from head movement patterns, our classification accuracy values were only moderately high. Assuming that our feature engineering approach was appropriate for capturing developmental effects, this performance gap may support the hypothesis that nonverbal communication behavior stabilizes at a certain age, after which it does not show significant developmental variation [24]. This hypothesis was also supported by the fact that our regression model was relatively less accurate; classifying individuals based on a developmentally informed age threshold proved to be more robust than predicting exact chronological age. In other words, in our age group of school-aged children, adolescents, and adults, head movements may not exhibit the necessary variation to make such a dimensional (continuous) prediction. Future research aimed at identifying precisely when head movement patterns stabilize in development is needed to reach a more informed conclusion.

The most important features, determined by the classifier, showed increased or decreased occurrences of head movements in different directions. Increased occurrences in the younger age group relative to the older age group were more dominant in the *monadic* case. In the *dyadic* case, both age groups were equally represented in terms of most important features. Notably, the features incorporating data from confederates were significantly more determining. It should be noted that our features measure movements in individual angles (*i.e.*, yaw, roll, or pitch); however, conversational

head movements usually involve movement in multiple angles simultaneously. Therefore, more sophisticated features that capture movements with respect to multiple angles should be considered in future work.

There are several limitations to the current work that need to be addressed in future studies. First, our multivariate analysis (*i.e.*, predicting age) is an indirect approach for studying how head movements change with age. A univariate approach of correlating individual head movement variables with age would be more direct. In this case, the number of possible features that needed to be included in the analysis (*e.g.*, 48 in *monadic* and 432 in *dyadic* cases) prevented a meaningful univariate study, making our multivariate approach the natural choice. This approach has also been widely used in neuroscience studies, in so-called “brain age” works [40].

Second, our treatment of the *dyadic* features does not capture *coordination* between the participant and the confederate. A proper coordination metric must capture simultaneous (and with possible lags) actions of the partners, a property that our *dyadic* features lack. Inclusion of such a metric may yield even higher accuracy, suggesting that how head movements are coordinated between partners, in particular, changes with age.

Third, our results are conclusive to the extent that our computational pipelines (feature construction and machine learning) are appropriate for capturing developmental effects. Exploring more sophisticated feature structures and machine learning models, such as the use of neural network architectures accounting for the time dependence of signals, is a viable next step. Similarly, generalizing our findings to different samples, such as additional age ranges or diverse cultural and ethnic groups, is currently difficult. Future

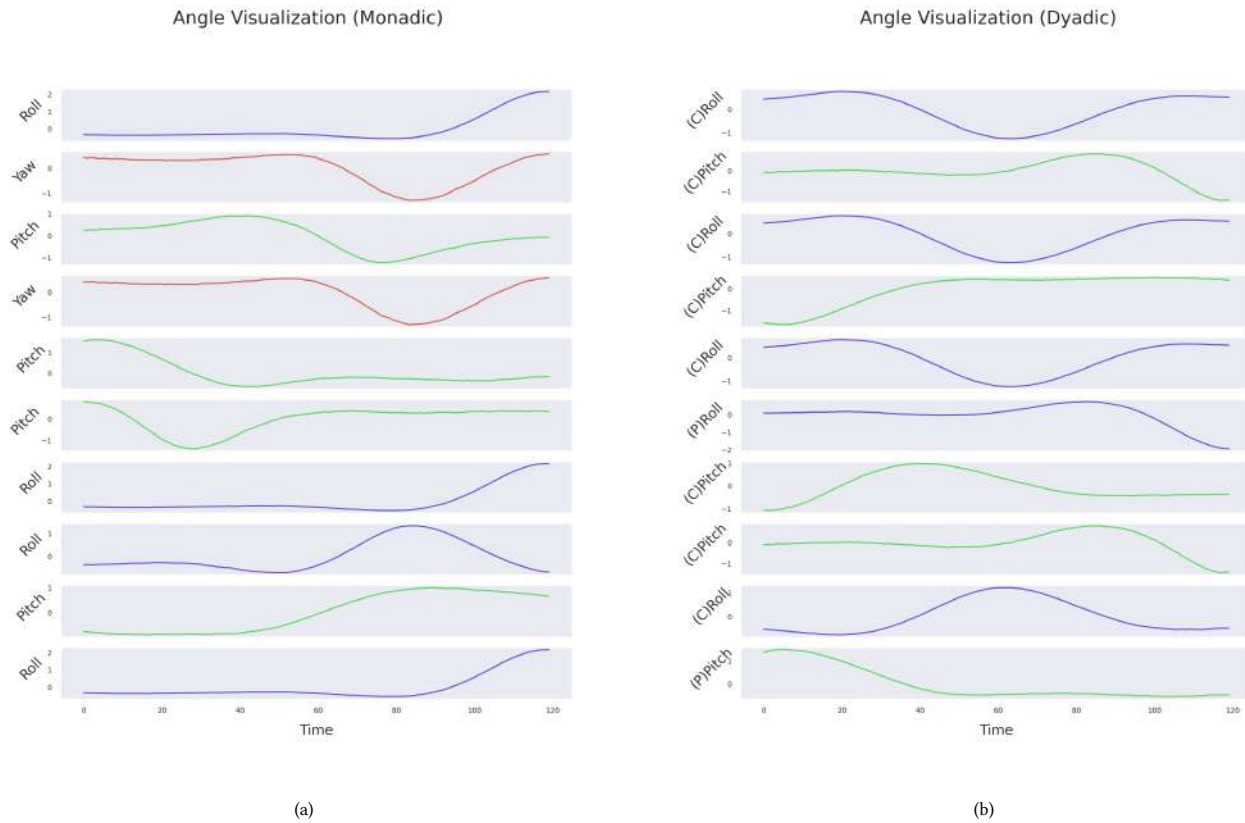


Figure 6: Time-dependent plots of features considered most important during SVM classification (top 10 features shown in Figure 5). In (a), head angle patterns were extracted using *monadic* features, and in (b), *dyadic* features. In each case, the cluster centers of head angles have been visualized. Each subplot shows the cluster center belonging to one angle. The x-axis shows the time dimension, and the y-axis, the change in the angle (normalized). With the *dyadic* features, the y-axis also contains information about who that feature belonged to: the participant (P) or confederate (C).

studies should consider using larger samples with more diverse characteristics to improve understanding of nonverbal communication variation across a wider range of demographic factors.

Our findings, in conjunction with future studies, will pave the way for a better understanding of developmentally-sensitive normative models and expected patterns of nonverbal communication. These models will allow us to chart individual differences relative to normative patterns. Understanding normative developmental trajectories in various domains of social communication behavior will also enhance conceptualization of mental health conditions in terms of deviations from those norms [28], facilitating individual-level analysis (*i.e.*, precision medicine approach) of observed phenotypes as well as underlying biological mechanisms.

6 CONCLUSION

This work focused on creating computational methods to extract meaningful information from head movement patterns, a particularly under-studied area of nonverbal communication. We aimed to examine whether head movement patterns during conversations

distinguished between individuals who were younger vs older than 12 years. Video data was collected from 3-minute face-to-face, semi-structured conversations, between each of 79 participants and a research confederate. The data were captured using our in-house ‘BioSensor’ system of two synchronized cameras, one facing the participant, and the other the confederate. Through unsupervised and supervised machine learning strategies, we achieved a classification accuracy of 78% relying on *dyadic* features, where the head movement patterns of both the participant and confederate were included in the analysis. A regression analysis confirmed the statistical significance of age prediction based on head movement patterns. In addition, we extracted head movement patterns that were relevant for the classification process, suggesting that they were potentially high in information content. These findings open doors for further social-behavioral, clinical, and human-computer interaction applications by demonstrating that nuanced nonverbal information extracted unobtrusively from short face-to-face conversations can have meaningful implications for both typical and atypical development.

ACKNOWLEDGMENTS

This work is partially funded by the National Institutes of Health (NIH) Office of the Director (OD) and National Institute of Mental Health (NIMH) of US, under grants R01MH122599, R01MH118327, 5P50HD105354-02, and 1R01MH125958-01.

REFERENCES

- [1] Kathleen T Ashenfelter, Steven M Boker, Jennifer R Waddell, and Nikolay Vitanov. 2009. Spatiotemporal symmetry and multifractal structure of head movements during dyadic conversation. *Journal of Experimental Psychology: Human Perception and Performance* 35, 4 (2009), 1072.
- [2] American Psychiatric Association et al. 2013. *Diagnostic and Statistical Manual of Mental Disorders (DSM-V)(5th ed.)*.
- [3] Judee K Burgoon, Valerie Manusov, and Laura K Guerrero. 2021. *Nonverbal communication*. Routledge.
- [4] Joseph N Cappella. 1981. Mutual influence in expressive behavior: Adult-adult and infant-adult dyadic interaction. *Psychological bulletin* 89, 1 (1981), 101.
- [5] David Curto, Albert Clapés, Javier Selva, Sorina Smeureanu, Julio Junior, CS Jacques, David Gallardo-Pujol, Georgina Guilera, David Leiva, Thomas B Moeslund, et al. 2021. Dyadformer: A multi-modal transformer for long-range modeling of dyadic interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2177–2188.
- [6] Starkey Duncan. 1972. Some signals and rules for taking speaking turns in conversations. *Journal of personality and social psychology* 23, 2 (1972), 283.
- [7] Javier Fenollar-Cortés, Ana Gallego-Martínez, and Luis J Fuentes. 2017. The role of inattention and hyperactivity/impulsivity in the fine motor coordination in children with ADHD. *Research in developmental disabilities* 69 (2017), 77–84.
- [8] Paula Fitzpatrick, Jean A Frazier, David M Cochran, Teresa Mitchell, Caitlin Coleman, and et RC Schmidt. 2016. Impairments of social motor synchrony evident in autism spectrum disorder. *Frontiers in Psychology* 7 (2016), 1323.
- [9] M Franchini, E Duku, V Armstrong, J Brian, SE Bryson, N Garon, W Roberts, C Roncadin, L Zwaigenbaum, and IM Smith. 2018. Variability in verbal and nonverbal communication in infants at risk for autism spectrum disorder: Predictors and outcomes. *Journal of Autism and Developmental disorders* 48, 10 (2018), 3417–3431.
- [10] Alexandra Livia Georgescu, Jana Christina Koehler, Johanna Weiske, Kai Vogeley, Nikolaos Koutsouleris, and Christine Falter-Wagner. 2019. Machine learning to study social interaction difficulties in ASD. *Frontiers in Robotics and AI* 6 (2019), 132.
- [11] Sylvie Goldman, Cuiling Wang, Miran W Salgado, Paul E Greene, Mimi Kim, and Isabelle Rapin. 2009. Motor stereotypies in children with autism and other developmental disorders. *Developmental Medicine & Child Neurology* 51, 1 (2009), 30–38.
- [12] Karl Grammer, Kirsten B Kruck, and Magnus S Magnusson. 1998. The courtship dance: Patterns of nonverbal synchronization in opposite-sex encounters. *Journal of Nonverbal behavior* 22, 1 (1998), 3–29.
- [13] Charline Grossard, Laurence Chaby, Stéphanie Hun, Hugues Pellerin, Jérémy Bourgeois, Arnaud Dapogny, Huaxiong Ding, Sylvie Serret, Pierre Foulon, Mohamed Chetouani, et al. 2018. Children facial expression production: influence of age, gender, emotion subtype, elicitation condition and culture. *Frontiers in psychology* (2018), 446.
- [14] Raquel E Gur, Christian G Kohler, J Daniel Ragland, Steven J Siegel, Kathleen Lesko, Warren B Bilker, and Ruben C Gur. 2006. Flat affect in schizophrenia: relation to emotion processing and neurocognitive measures. *Schizophrenia bulletin* 32, 2 (2006), 279–287.
- [15] Joanna Hale, Jamie A Ward, Francesco Buccheri, Dominic Oliver, and Antonia F de C Hamilton. 2020. Are you on my wavelength? Interpersonal coordination in dyadic conversations. *Journal of nonverbal behavior* 44, 1 (2020), 63–83.
- [16] Judith A Hall, Terrence G Horgan, and Nora A Murphy. 2019. Nonverbal communication. *Annual review of psychology* 70, 1 (2019), 271–294.
- [17] Zakia Hammal, Jeffrey F Cohn, and David T George. 2014. Interpersonal coordination of headmotion in distressed couples. *IEEE transactions on affective computing* 5, 2 (2014), 155–167.
- [18] Zakia Hammal, Jeffrey F Cohn, and Daniel S Messinger. 2015. Head movement dynamics during play and perturbed mother-infant interaction. *IEEE transactions on affective computing* 6, 4 (2015), 361–370.
- [19] JingGuang Han, Nick Campbell, Kristiina Jokinen, and Graham Wilcock. 2012. Investigating the use of non-verbal cues in human-robot interaction with a Nao robot. In *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 679–683.
- [20] Mohammad Shahidul Islam and Ksenia Kirillova. 2020. Non-verbal communication in hospitality: At the intersection of religion and gender. *International Journal of Hospitality Management* 84 (2020), 102326.
- [21] Thomas Jahn, Rudolf Cohen, Werner Hubmann, Fritz Mohr, Iris Köhler, Regine Schlenker, Rainer Niethammer, and Johannes Schröder. 2006. The Brief Motor Scale (BMS) for the assessment of motor soft signs in schizophrenic psychoses and other psychiatric disorders. *Psychiatry research* 142, 2-3 (2006), 177–189.
- [22] Hifza Javed, WonHyong Lee, and Chung Hyuk Park. 2020. Toward an automated measure of social engagement for children with autism spectrum disorder—a personalized computational modeling approach. *Frontiers in Robotics and AI* (2020), 43.
- [23] Rana el Kaliouby and Peter Robinson. 2005. Real-time inference of complex mental states from facial expressions and head gestures. In *Real-time vision for human-computer interaction*. Springer, 181–200.
- [24] Erin A Koterba. 2010. *Conversations between friends: Age and context differences in the development of nonverbal communication in preadolescence*. University of Pittsburgh.
- [25] Pradeep Raj Krishnappa Babu, J Matias Di Martino, Zhuoqing Chang, Sam Perchon, Rachel Aiello, Kimberly LH Carpenter, Scott Compton, Naomi Davis, Lauren Franz, Steven Espinosa, et al. 2022. Complexity analysis of head movements in autistic toddlers. *Journal of Child Psychology and Psychiatry* (2022).
- [26] Zeno Kupper, Fabian Ramseyer, Holger Hoffmann, and Wolfgang Tschacher. 2015. Nonverbal synchrony in social interactions of patients with schizophrenia indicates socio-communicative deficits. *PLoS One* 10, 12 (2015), e0145882.
- [27] Steven R Livingstone and Caroline Palmer. 2016. Head movements encode emotions during speech and song. *Emotion* 16, 3 (2016), 365.
- [28] Andre F Marquand, Seyed Mostafa Kia, Mariam Zabihi, Thomas Wolfers, Jan K Buitelaar, and Christian F Beckmann. 2019. Conceptualizing mental disorders as deviations from normative functioning. *Molecular psychiatry* 24, 10 (2019), 1415–1424.
- [29] Katherine B Martin, Zakia Hammal, Gang Ren, Jeffrey F Cohn, Justine Cassell, Mitsunori Ogihara, Jennifer C Britton, Anibal Gutierrez, and Daniel S Messinger. 2018. Objective measurement of head movement differences in children with and without autism spectrum disorder. *Molecular autism* 9, 1 (2018), 1–10.
- [30] Evelyn Z McClave. 2000. Linguistic functions of head movements in the context of speech. *Journal of pragmatics* 32, 7 (2000), 855–878.
- [31] Raveesh Meena, Kristiina Jokinen, and Graham Wilcock. 2012. Integration of gestures and speech in human-robot interaction. In *2012 IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE, 673–678.
- [32] Krysta Murray, Linda Lillakas, Rebecca Weber, Suzanne Moore, and Elizabeth Irving. 2007. Development of head movement propensity in 4–15 year old children in response to visual step stimuli. *Experimental brain research* 177, 1 (2007), 15–20.
- [33] Julia Parish-Morris, Evangelos Sariyanidi, Casey Zampella, G Keith Bartley, Emily Ferguson, Ashley A Pallathra, Leila Bateman, Samantha Plate, Meredith Cola, Juhi Pandey, et al. 2018. Oral-Motor and lexical diversity during naturalistic conversations in adults with autism spectrum disorder. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, Vol. 2018. NIH Public Access, 147.
- [34] Deepika Phutela. 2015. The importance of non-verbal communication. *IUP Journal of Soft Skills* 9, 4 (2015), 43.
- [35] Allison B Ratto, Lauren Turner-Brown, Betty M Rupp, Gary B Mesibov, and David L Penn. 2011. Development of the contextual assessment of social skills (CASS): A role play measure of social skill for individuals with high-functioning autism. *Journal of autism and developmental disorders* 41, 9 (2011), 1277–1286.
- [36] Evangelos Sariyanidi, Casey J Zampella, Robert T Schultz, and Birkan Tunç. 2020. Inequality-constrained and robust 3D face model fitting. In *European Conference on Computer Vision*. Springer, 433–449.
- [37] Lars Schillingmann, Joseph M Burling, Hanako Yoshida, and Yukie Nagai. 2015. Gaze is not enough: Computational analysis of infant’s head movement measures the developing response to social interaction. (2015).
- [38] Edgar Seemann, Kai Nickel, and Rainer Stiefelwagen. 2004. Head pose estimation using stereo vision for human-robot interaction. In *Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings. IEEE*, 626–631.
- [39] Ullica Segerstråle and Peter Molnár. 2018. *Nonverbal communication: where nature meets culture*. Routledge.
- [40] Birkan Tunç, Lisa D Yankowitz, Drew Parker, Jacob A Alappatt, Juhi Pandey, Robert T Schultz, and Ragini Verma. 2019. Deviation from normative brain development is associated with symptom severity in autism spectrum disorder. *Molecular autism* 10, 1 (2019), 1–14.
- [41] Manuel Varlet, Ludovic Marin, Delphine Capdevielle, Jonathan Del-Monte, Richard C Schmidt, Robin N Salese, Jean-Philippe Boulenger, Benoît G Bardy, and Stéphane Raffard. 2014. Difficulty leading interpersonal coordination: towards an embodied signature of social anxiety disorder. *Frontiers in Behavioral Neuroscience* 8 (2014), 29.
- [42] Akiko Yamazaki, Keiichi Yamazaki, Yoshinori Kuno, Matthew Burdelski, Michie Kawashima, and Hideaki Kuzuoka. 2008. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 131–140.
- [43] Casey J Zampella, Evangelos Sariyanidi, Anne G Hutchinson, G Keith Bartley, Robert T Schultz, and Birkan Tunç. 2021. Computational Measurement of Motor Imitation and Imitative Learning Differences in Autism Spectrum Disorder.

In *Companion Publication of the 2021 International Conference on Multimodal Interaction*. 362–370.

- [44] Casey J Zampella, Leah AL Wang, Margaret Haley, Anne G Hutchinson, and Ashley de Marchena. 2021. Motor skill differences in autism spectrum disorder: A clinically focused review. *Current psychiatry reports* 23, 10 (2021), 1–11.
- [45] Zhong Zhao, Zhipeng Zhu, Xiaobin Zhang, Haiming Tang, Jiayi Xing, Xinyao Hu, Jianping Lu, Qiongleng Peng, and Xingda Qu. 2021. Atypical Head Movement

during Face-to-Face Interaction in Children with Autism Spectrum Disorder. *Autism Research* 14, 6 (2021), 1197–1208.

- [46] Zhong Zhao, Zhipeng Zhu, Xiaobin Zhang, Haiming Tang, Jiayi Xing, Xinyao Hu, Jianping Lu, and Xingda Qu. 2022. Identifying autism with head movement features by implementing machine learning algorithms. *Journal of Autism and Developmental Disorders* 52, 7 (2022), 3038–3049.