Published in final edited form as:
Proc ACM Int Conf Multimodal Interact. 2024 November ; 2024: 350–354. doi:10.1145/3678957.3685711.

Detecting Autism from Head Movements using Kinesics

Muhittin Gokmen,
MEF University

Evangelos Sariyanidi,
Children's Hospital of Philadelphia

Lisa Yankowitz,
Children's Hospital of Philadelphia

Casey J. Zampella,
Children's Hospital of Philadelphia

Robert T. Schultz,
Children's Hospital of Philadelphia, University of Pennsylvania

Birkan Tunç
Children's Hospital of Philadelphia, University of Pennsylvania

Abstract

Head movements play a crucial role in social interactions. The quantification of communicative movements such as nodding, shaking, orienting, and backchanneling is significant in behavioral and mental health research. However, automated localization of such head movements within videos remains challenging in computer vision due to their arbitrary start and end times, durations, and frequencies. In this work, we introduce a novel and efficient coding system for head movements, grounded in Birdwhistell's kinesics theory, to automatically identify basic head motion units such as nodding and shaking. Our approach first defines the smallest unit of head movement, termed *kine*, based on the anatomical constraints of the neck and head. We then quantify the location, magnitude, and duration of *kines* within each angular component of head movement. Through defining possible combinations of identified *kines*, we define a higher-level construct, *kineme*, which corresponds to basic head motion units such as nodding and shaking. We validate the proposed framework by predicting autism spectrum disorder (ASD) diagnosis from video recordings of interacting partners. We show that the multi-scale property of the proposed framework provides a significant advantage, as collapsing behavior across temporal scales reduces performance consistently. Finally, we incorporate another fundamental behavioral modality, namely speech, and show that distinguishing between speaking- and listening-time head movementsments significantly improves ASD classification performance.

gokmenm@mef.edu.tr .

Keywords

Head Movements; Kinesics; Computer Vision; Psychology; Autism

1 INTRODUCTION

Head movements are an integral element of social communication, and an increasing body of research highlights their importance for characterizing social behavior in behavioral sciences and mental health. In autism, a neurodevelopment condition characterized in part by difficulties in social communications, alterations in head movements have been documented repeatedly [6, 8], motivating researchers to use computational approaches for studying them in various contexts (Section 2).

Analyzing head motion presents several challenges. The duration, speed, and other kinematic properties of head movements can vary greatly based on many factors including individual differences, social contexts, and changing communicative intent. A promising approach is breaking down movement signals into basic building blocks, similar to phonemes in language [22], action units in facial expressions (AUs) [4], or hand muscle action units in gestures [9]. Such coding systems can distinguish between the form, duration, and magnitude of basic movements. However, to our knowledge, there is no commonly agreed coding system for head movements.

In this work, we introduce a novel framework that operationalizes the kinesics theory pioneered by Birdwhistell [2] to create and implement a computerized coding system for analyzing head movements, allowing researchers to automatically detect the form, duration, and intensity of movements. Unlike previous computational works on kinesics, our approach, for the first time implements the theoretical framework by Birdwhistell with knowledge- rather than data-driven approach (Section 2), therefore the proposed features do not change from study to study and can serve as a common coding system. We first propose a method to automatically detect the location, magnitude, and scale of *kine*s, the smallest unit of head movement, within angular time series of head movements, based on the anatomical constraints of the neck and head. Next, a higher-level construct, *kineme*, is defined by combining *kines* in all possible permutations at multiple scales. The *kinemes* defined in this way correspond to basic head motion units such as nodding and shaking. We validate our framework by showing that it can capture meaningful head movement differences in autism spectrum disorder (ASD), and investigate the benefits of integrating another modality, namely speech, to separately analyze speaking- and listening-time behavior.

The contributions of this study are threefold. First, we propose the first knowledge-driven implementation of the kinesics framework that can detect basic movement patterns at different magnitudes, intensities, and time scales. Second, we show that using speech modality by distinguishing between speaking- and listening-time behavior leads to increased ASD classification performance. Finally, we show that our multi-scale approach leads to consistent performance improvement compared to a scale-agnostic approach.

2 RELATED WORK

Previous works used computer vision to study alterations in head movements [10] during different behavioral tasks such as watching videos [14], response to name [3, 16], or face-to-face interactions [15, 18, 24]. The features used in these studies are either ad-hoc [24], or based on classical signal processing techniques such as k-means clustering and bag-of-words [15], or cross-correlation [18], which are useful for classification but can be limited in terms of explainability [5]. Kinesics is a conceptual framework [2] that offers an alternative approach by breaking movements down into simpler patterns, which facilitates the investigation of the movement patterns that are most informative. As such, computer scientists were motivated to operationalize this conceptual framework. Xiao et al. [23] identified head motion patterns through a series of steps including computing the optical flow, segmenting motion patterns, representing these segments via Line Spectral Frequencies, and clustering them. Upon adopting a similar approach, we encountered limitations in achieving homogenous movement patterns. Consequently, our focus shifted towards extracting more robust and smaller elementary units capable of handling various magnitudes and duration of head motion patterns. Gahalawat et al. [5] defined *kinemes* as learned features, and showed that they can achieve state-of-the-art performance in depression classification. However, since learned features are study-dependent, they pose limitations in defining a common coding system. Thus, we propose a framework based on the first principles imposed by the anatomical restrictions (Section 3.1). Further, we propose the first multi-scale kinesics framework, and experiments show that this is a critical advantage as opposed to a scale-agnostic approach (Section 5).

3 HEAD MOVEMENT ANALYSIS VIA KINESICS

Kinesics represents a systematic study of visually perceptible aspects of nonverbal communication [1, 2]. Kinesics proposes a hierarchical structure for the body motion similar to the structure of language (*e.g.*, phones, phonemes), yielding terms like kine and kineme. A kine represents the isolable fundamental unit, while *kineme*, akin to phonemes in spoken language, is the most elementary and meaningful unit of motion, such as a full head nodding.

3.1 Identifying *Kines*

Given time series data corresponding to pitch, yaw, and roll angles of head motion, estimated using computer vision algorithms [19], the smallest structural elements within these time series can be identified by exploiting an anatomical constraint. When the head moves in any direction through the contraction of specific neck muscles, it tends to return to the resting position. This movement generates angles that increase (or decrease) in magnitude to a maximum value and then decrease (or increase) to return to the initial value within a certain duration, manifesting as a peak or a valley. Thus, by detecting peaks and valleys across different time scales, *kines* can be detected automatically. Various multi-scale methods, including Scale Invariant Feature Transform (SIFT) [11, 12], wavelets [13, 20] and Convolutional Neural Networks (CNN) [20], can be employed. In this study, we utilized a 1D SIFT to determine the local maxima and minima in the Laplacian of Gaussian (LoG) pyramid of the time series in both time and scale spaces, yielding a list of *kines* (peak or valley) with parameters including position, magnitude, and scale (duration), as illustrated in Figure 1a.

3.2 Identifying *Kinemes*

For each time series of head motion (pitch, yaw, or roll), there are three possible symbols, namely peak (P), valley (V), or null (N) at a given time frame. Thus, 3^3 = 27 possible combinations exist for the three time series (*e.g.*, PPP, PVP, PNV, NNN). *Kinemes* are defined by assigning a letter for each combination, except the null combination (NNN), spanning all 26 letters of the English alphabet (*e.g.*, PPP: a, PPV: b, PPN: c, PVP: d). Among these 26 letters, six are called “singletons”, including a non-null value only in one of the time series (*i.e.*, PNN, NPN, NNP, VNN, NVN, NNV). For example, *kineme* “i” (PNN) corresponds to the head moving up and down once, while *kineme* “r” (VNN) corresponds to the down and up movement.(Code is available at: https:/github.com/gokmenm/hma_kinesics.)

To determine the *kineme* code from real-valued *kines*, we utilize the *kineme cube*, where each *kineme* is positioned at a corner or midpoint on the face of a unit cube. Each *kineme* is represented as a unit vector from the origin (0, 0, 0) to one of these corners and midpoints, resulting in 26 unit vectors. The real-valued *kine* vector of the actual data is constructed as the vector, $\mathbf{x} = [m_{pitch}, m_{yaw}, m_{roll}]^T$, where m_{pitch}, m_{yaw}, and m_{roll} are LoG responses (normalized to have a unit vector) for pitch, roll, and pitch components respectively. We then compute the cosine distance between the vector **x** and 26 *kineme* vectors, and assign the *kineme* with the lowest distance. When repeated for all the frames of a given video recording, this procedure yields a sequence of letters, such as "——x——ix——-uuif—ci-r–lr-pc-rfx–xr-cru-irl-fxior-", where “-” is the null character.

3.3 A Head Motion Production Model

Similar to the speech production model, we can reconstruct the original head motion signal using the detected *kines*. For clarity, we define the reconstruction model for a single angle here. The same procedure can be applied to the other two angles. Let $k = (t, m, \sigma)$ defines a *kine* with position, t, magnitude, m, and scale, σ. As a head motion generated at time t is smoothed out by the muscle controlling the movement, a *kine* can be modeled as an impulse at time t, smoothed by a low-pass filter as follows:

$$k = (\sigma / \sqrt{2\pi})m\,(\delta(t) * G(t, \sigma)) + N(0, \sigma_n), \tag{1}$$

where m is the magnitude ($m > 0$ for a peak and $m < 0$ for a valley), $\delta(t)$ is the Dirac delta function representing the impulse at t, $G(t, \sigma)$ is Gaussian filter with a standard deviation σ, and $N(0, \sigma_n)$ is a zero mean noise with standard deviation σ_n, and * denotes the convolution operation. Figure 1b shows an example reconstruction of the original pitch angle from detected *kines*.

4 ASD CLASSIFICATION USING KINESICS

We derive kineme histograms to perform ASD classification (Section 4.1) and use automated speech detection to distinguish between listening- and speaking-time behavior (Section 4.2).

4.1 Kineme histograms as head motion features

The distribution of *kinemes* over a selected time interval provides a summary of dyadic communication. We quantify kineme distributions through histograms of counts or magnitudes of *kinemes*, extracted at the four scales (levels 1,4,7 and 10 of SIFT). The videos are split into non-overlapping time segments, then separate histograms are extracted per segment, and then they are concatenated. We tested the contribution of different types of *kinemes* by generating histograms including only singleton letters, or including all letters. We aggregate kinemes from different scales in different histograms, and then concatenate all histograms. For comparison, we also aggregate kinemes from all scales in a single histogram.

4.2 Separating speaker and listener behavior

Since the head movements of a speaker and listener typically convey different meanings, we used automated speech detection to investigate the effect of distinguishing between speaking- and listening-time behavior. We compared an audio-visual speech detection algorithm [21] with a CNN-based visual voice activity detector that we trained using public data [7], and used the latter as it is simpler yet did not provide worse results in our preliminary experiments.

5 EXPERIMENTS

We conducted experiments on two datasets of different age ranges. We validated performance in terms of ASD vs. neurotypical (NT) classification; performed an ablation study to investigate the effect of the fundamental design choices in the kineme histograms; and showed how the proposed method can be used for investigating specific research questions (Section 5.1).

Datasets.

Data was collected at the Children’s Hospital of Philadelphia, and the research was approved by its institutional review board. This study included two samples, namely adolescents (ages 12-17 years) and young adults (ages 19-49). The young adult sample included 16 ASD and 27 NT participants, and the adolescent sample included 26 NT and 38 ASD participants (Table 1).

The experimental procedure in both datasets consisted of a semi-structured assessment of conversational ability designed to mimic real-life first-time encounters [17]. Participants engaged in a 3-5 minute face-to-face conversation (Table 1) with a research confederate while frontal face videos of both the participant and confederate were captured [15]. Head pose data was extracted using 3DI[19]. The output consisted of time-dependent signals for the three head movement angles: pitch (head nodding axis), yaw (head shaking axis), and roll (tilting axis). In this study, time series of the participant alone were used for the analysis. It must be noted that some characteristics of this sample (*e.g.*, American, English-speaking) may impact the frequency of specific head movements.

Experimental setup and baselines.

We compare the proposed kineme histograms with two standard alternative features for head movements, namely K-means (bag-of-words) [15] and intra-person windowed cross-correlation (WCC) [18]. Results for our method and WCC are obtained via nested cross-validation across the two datasets (i.e., adolescents and young adults). Specifically, we obtain the 10 best-performing hyperparameter combinations for each feature on one dataset and then apply them through nested leave-one-out cross validation on the other (i.e., one of the ten hyperparameter combinations is selected at each outer fold). Such a cross-dataset approach guarantees the generalizability of our findings. The hyperparameters tested for WCC are time window size (1s, 2s, 4s, 6s), allowed lag (0.5s, 1s, 2s) and step size (1s, 2s); and for our kineme approach are histogram type (count or magnitude), treatment of scales (separated or collapsed), and histogram length (6s, 10s, 20s). The result for the K-means approach is taken from [15], where the young adults dataset was used. When comparing with other methods, for a fair comparison, our results were computed without speech detection. The classifier used for all features is linear SVM.

5.1 Results

Table 2 shows results with the three different feature types. *Kineme* results are reported for two cases: (i) when all letters are used, and (ii) when only singletons are used. WCC features perform well on the young adult dataset but not on the adolescent dataset. On the other hand, *kinemes* from all letters perform well on the adolescents dataset but not on the young adult dataset. *Kineme* histograms from singletons yield the best result on both datasets, indicating that singletons are likely more informative than the remaining letters.

Ablation study: Effect of design choices.—Table 3 provides classification results with different design choices on a combined dataset that includes both the young adult sample and the adolescent sample. Results indicate a number of clear trends. First, separating scales into different histograms is always better than collapsing them into a single histogram. Second, listening behavior alone yields consistently lower classification accuracy than speaking behavior. The best results are obtained by generating different histograms for speaking and listening states and using them both. This approach can capture possible differences in ASD related to the integration of different behavioral modalities (*i.e.*, speech and head movements). Finally, consistently with results in Table 2, the usage of singletons yields better results than using all the letters.

Investigating specific head movements.—An advantage of the kinesics approach is that one can conduct scientific analyses that investigate the effect of specific head movements by restricting the study to the corresponding letters. We demonstrate such a use case by analyzing the ability of two fundamental head movements, namely nodding and shaking, to predict ASD. Figure 2 shows classification results using only the nod- or the shake-related kinemes by visualizing the distance to SVM hyperplane and classification accuracy. First, results further highlight the advantage of the multiscale kinesics approach, as identifying letters across all scales and treating those scales independently (top row) leads to significantly better results than collapsing across scales (bottom row). Further, the classifier based on nodding alone achieves significantly higher accuracy than the one based on shaking, indicating that the former movement likely carries richer information.

6 CONCLUSION

We provided a novel, theoretical framework for implementing the kinesics approach of Birdwhistell [2]. Our framework detects basic head movements in forms of *kinemes*, and quantifies their temporal scale as well as magnitude. Experiments show that the proposed *kineme*s can successfully distinguish the head movements of individuals with autism from those of neurotypical individuals. In particular, the multi-scale property of our framework leads to significant improvement in classification accuracy. Encoding speaking- and listening-time behavior separately increases accuracy further.

ACKNOWLEDGMENTS

The work of M. Gokmen is partially supported by the BIDEB2219 program of the Scientific and Technological Research Council of Turkey (TUBITAK) under the grant # 1059B192300279. The work of the other co-authors is partially supported by the Office of the Director (OD), National Institute of Child Health and Human Development (NICHD), and National Institute of Mental Health (NIMH) of US, under grants R01MH118327, R01MH122599, 5P50HD105354-02 and R21HD102078; and the IDDRC at CHOP/Penn.

REFERENCES

[1]. Birdwhistell Ray L.. 1952. Introduction to kinesics: an annotation system for analysis of body motion and gesture. Reprint, University of Michigan Press, Ann Arbor 2021.

[2]. Birdwhistell Ray L.. 1970. Kinesics and Context, Essays on Body Motion Communication. University of Pennsylvania Press.

[3]. Campbell Kathleen, Carpenter Kimberly L.H., Hashemi Jordan, Espinosa Steven, Marsan Samuel, Borg Jana Schaich, Chang Zhuoqing, Qiu Qiang, Vermeer Saritha, Adler Elizabeth, Tepper Mariano, Egger Helen L., Baker Jeffery P., Sapiro Guillermo, and Dawson Geraldine. 2019. Computer vision analysis captures atypical attention in toddlers with autism. Autism 23 (2019), 619–628. Issue 3. 10.1177/1362361318766247 [PubMed: 29595333]

[4]. Friesen E and Ekman Paul. 1978. Facial action coding system: a technique for the measurement of facial movement. Palo Alto 3, 2 (1978), 5.

[5]. Gahalawat Monika, Rojas Raul Fernandez, Guha Tanaya, Subramanian Ramanathan, and Goecke Roland. 2023. Explainable Depression Detection via Head Motion Patterns. In Proceedings of the 25th International Conference on Multimodal Interaction (, Paris, France,) (ICMI '23). Association for Computing Machinery, New York, NY, USA, 261–270. 10.1145/3577190.3614130

[6]. Goldman Sylvie, Wang Cuiling, Salgado Miran W., Greene Paul E., Kim Mimi, and Rapin Isabelle. 2009. Motor stereotypies in children with autism and other developmental disorders. Developmental Medicine and Child Neurology 51 (2009), 30–38. Issue 1. 10.1111/j.1469-8749.2008.03178.x

[7]. Guy Sylvain, Lathuilière Stéphane, Mesejo Pablo, and Horaud Radu. 2021. Learning visual voice activity detection with an automatically annotated dataset. In 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 4851–4856.

[8]. Hutt Corinne and Hutt SJ. 1965. Effects of environmental complexity on stereotyped behaviours of children. Animal Behaviour 13 (1965), 1–4. Issue 1. 10.1016/0003-3472(65)90064-3

[9]. Ho-Shing Ip Horace, Chan Sam CS, and Lam Maria SW. 1998. HACS: Hand Action Coding System for anatomy-based synthesis of hand gestures. In SMC'98 Conference Proceedings. 1998 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 98CH36218), Vol. 2. IEEE, 1207–1212.

[10]. Koehler Jana Christina, Dong Mark Sen, Bierlich Afton M., Fischer Stefanie, Späth Johanna, Plank Irene Sophia, Koutsouleris Nikolaos, and Falter-Wagner Christine M.. 2024. Machine learning classification of autism spectrum disorder based on reciprocity in naturalistic social interactions. Translational Psychiatry 14 (2024), 76–84. Issue 1. 10.1038/s41398-024-02802-5 [PubMed: 38310111]

[11]. Lowe David G.. 1999. Object recognition from local scale-invariant features. ProcProceedings of the IEEE International Conference on Computer Vision 2, 1150–1157. 10.1109/iccv.1999.790410

[12]. Lowe David G.. 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60 (2004), 91–110. Issue 2. 10.1023/B:VISI.0000029664.99615.94

[13]. Mallat Stephane and Zhong Sifen. 1992. Characterization of Signals From Multiscale Edges. IEEE Transactions on Pattern Analysis and Machine Intelligence 14 (1992), 710–732. Issue 7. 10.1109/34.142909

[14]. Martin Katherine B., Hammal Zakia, Ren Gang, Cohn Jeffrey F., Cassell Justine, Ogihara Mitsunori, Britton Jennifer C., Gutierrez Anibal, and Messinger Daniel S.. 2018. Objective measurement of head movement differences in children with and without autism spectrum disorder. Molecular Autism 9 (2018), 14. Issue 1. 10.1186/s13229-018-0198-4 [PubMed: 29492241]

[15]. Denisa Qori McDonald Evangelos Sariyanidi, Zampella Casey J., Dejardin Ellis, Herrington John D., Schultz Robert T., and Tunc Birkan. 2023. Predicting Autism from Head Movement Patterns during Naturalistic Social Interactions. ACM International Conference Proceeding Series, 55–60. 10.1145/3608298.3608309

[16]. Perochon Sam, Matias Di Martino Rachel Aiello, Baker Jeffrey, Carpenter Kimberly, Chang Zhuoqing, Compton Scott, Davis Naomi, Eichner Brian, Espinosa Steven, Flowers Jacqueline, Franz Lauren, Gagliano Martha, Harris Adrianne, Howard Jill, Kollins Scott H., Perrin Eliana M., Raj Pradeep, Spanos Marina, Walter Barbara, Sapiro Guillermo, and Dawson Geraldine. 2021. A scalable computational approach to assessing response to name in toddlers with autism. Journal of Child Psychology and Psychiatry and Allied Disciplines 62 (2021), 1120–1131. Issue 9. 10.1111/jcpp.13381 [PubMed: 33641216]

[17]. Ratto Allison B., Turner-Brown Lauren, Rupp Betty M., Mesibov Gary B., and Penn David L.. 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 (2011), 1277–1286. Issue 9. 10.1007/s10803-010-1147-z [PubMed: 21287253]

[18]. Sariyanidi Evangelos, Zampella Casey J, DeJardin Ellis, Herrington John D, Schultz Robert T., and Tunc Birkan. 2023. Comparison of Human Experts and AI in Predicting Autism from Facial Behavior. In CEUR workshop proceedings, Vol. 3359. NIH Public Access, 48.

[19]. Sariyanidi Evangelos, Zampella Casey J., Schultz RobertT., and Tunc Birkan. 2024. Inequality-Constrained 3D Morphable Face Model Fitting. IEEE Transactions on Pattern Analysis and Machine Intelligence 46 (2024), 1305–1318. Issue 2. 10.1109/TPAMI.2023.3334948 [PubMed: 38015704]

[20]. Schmidt Mikkel N., Alstrom Tommy S., Svendstorp Marcus, and Larsen Jan. 2019. Peak Detection and Baseline Correction Using a Convolutional Neural Network. Proc. IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP 2019-May, 2757–2761. 10.1109/ICASSP.2019.8682311

[21]. Tao Ruijie, Pan Zexu, Das Rohan Kumar, Qian Xinyuan, Shou Mike Zheng, and Li Haizhou. 2021. Is Someone Speaking?: Exploring Long-term Temporal Features for Audio-visual Active Speaker Detection. MM 2021 - Proceedings of the 29th ACM International Conference on Multimedia, 3927–3935. 10.1145/3474085.3475587

[22]. Wells John C. 1982. Accents of English: Volume 1. Vol. 1. Cambridge University Press.

[23]. Xiao Bo, Georgiou Panayiotis, Baucom Brian, and Narayanan Shrikanth S.. 2015. Head motion modeling for human behavior analysis in dyadic interaction. IEEE Transactions on Multimedia 17 (7 2015), 1107–1119. Issue 7. 10.1109/TMM.2015.2432671 [PubMed: 26557047]

[24]. Zhao Zhong, Zhu Zhipeng, Zhang Xiaobin, Tang Haiming, Xing Jiayi, Hu Xinyao, Lu Jianping, and Qu Xingda. 2022. Identifying Autism with Head Movement Features by Implementing Machine Learning Algorithms. Journal of Autism and Developmental Disorders 52 (2022), 3038–3049. Issue 7. 10.1007/s10803-021-05179-2 [PubMed: 34250557]

CCS CONCEPTS

• Applied computing → Psychology;•Computing methodologies → Computer vision.

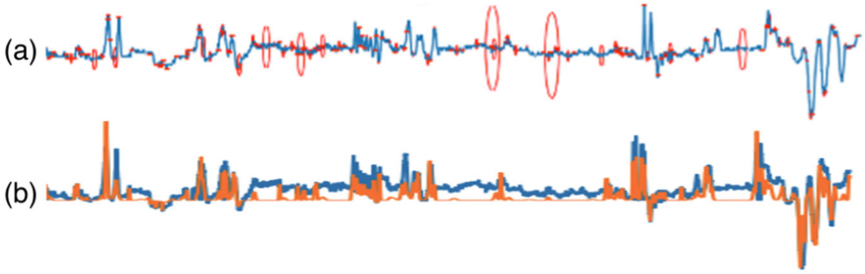

Figure 1:

(a) Detected *kine*s as peaks and valleys of the pitch angle of head rotation, where red points represents detected *kines*, while the size of red ellipse represents their scales. (b) Reconstructed (orange) signal and the original (blue) signal.

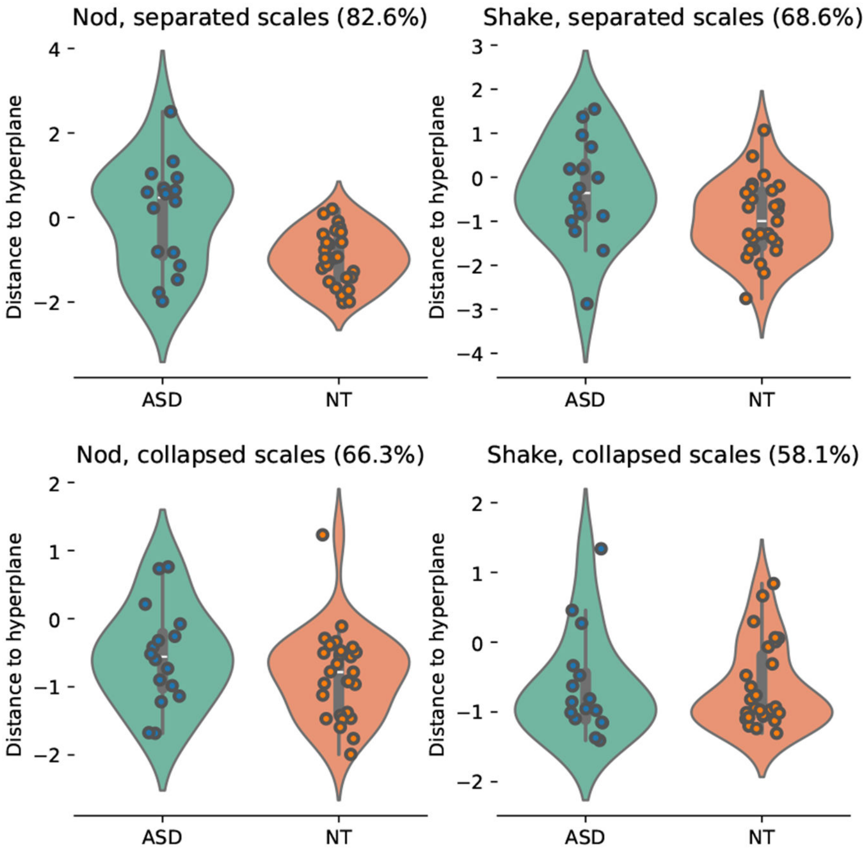

Figure 2:

Classification results on the young adult when only nodding or shaking kinemes are used. Results are reported separately for classifiers that are based on separated scales (top row) and classifiers that collapse kinemes across scales (bottom row). All classifiers use speech activity detection to distinguish between speaking- and listening-time behavior. Values within parentheses indicate classification accuracy.

Table 1:

The (mean) age, IQ and conversation duration; and number of female and male participants in the two study samples. *p* values indicate possible group differences.

	Adolescents			**Young adults**		
	ASD	**TDC**	***p* val.**	**ASD**	**TDC**	***p* val.**
Age	14.8	14.2	0.124	28.3	28.1	0.919
IQ	97.3	109	0.006	100	112	0.040
Duration (secs)	212	238	0.076	185	190	0.181
F/M participants	8/30	13/13	0.031	2/14	4/23	1.000

Table 2:

ASD vs. NT classification accuracy (%) on two datasets. Numbers in parentheses are balanced accuracy. Bold, italic and underlined texts respectively denote the first, second, and third best result.

	Young adults	**Adolescents**
Kinemes (Singletons)	**76.7** (75.1)	**79.7** (79.9)
Kinemes (All letters)	62.8 (61.0)	*74.6* (74.0)
WCC	*74.4* (74.0)	<u>60.7</u> (59.0)
K-means	<u>66.7</u>	N/A

Table 3:

Classification accuracy (%) on the combined dataset (i.e., adolescents and young adults) with kineme histograms. The effect of four factors is shown: Histogram type (count vs. magnitude), treatment of scales (separated or collapsed), letters used (singletons vs. all letters) and usage of speech detection (speech- and listening-time kinemes in different histograms combined, speech-time kinemes only, listening-time kinemes only, and no usage of speech separation). Bold, italic and underlined texts respectively denote the first, second, and third best result. †Spe: Speaking, Lis: Listening, Non-sep: Non-separated speech.

		Count Histograms				Magnitude Histograms			
		Spe & Lis	**Spe**	**Lis**	**Non-sep.**	**Spe & Lis**	**Spe**	**Lis**	**Non-sep.**
Separated scales	Singletons	*75.5*	71.6	66.7	*75.5*	**80.4**	71.6	53.9	63.7
	All letters	72.5	67.6	61.8	68.6	70.6	64.7	59.8	66.7
Collapsed scales	Singletons	61.8	68.6	60.8	57.8	64.7	56.9	48.0	52.9
	All letters	59.8	64.7	53.9	62.7	57.8	57.8	48.0	52.9