Blink as you Sync - Uncovering Eye and Nod Synchrony in Conversation using Wearable Sensing

Aman Gupta, Finn L Strivens*
Benjamin Tag, Kai Kunze
Graduate School of Media Design
Keio University
Yokohama, Japan
aman@keio.jp

Jamie A Ward
Department of Computing
Goldsmiths University of London
London, United Kingdom
jamie@jamieward.net

ABSTRACT

We tend to synchronize our movements to the person we are talking to during face-to-face conversation. Higher interpersonal synchrony is linked to greater empathy and more effortless interactions. This paper presents a first method and a corresponding dataset to explore synchrony in natural conversation by capturing eye and head movement using commodity smart eyewear. We present a 17 hour dataset, using Electrooculography and inertial sensing, of 42 people in conversation (21 dyads: 10 in Japanese, 10 in English, 1 in Chinese). Initial results on 18 dyads show significant interpersonal synchrony of blink and head nod behaviour during conversation (at frequencies of 0.2 to 0.5 Hz). We also find that people are more likely to synchronise blinks at around 1 Hz when conversing back-to-back than when face-to-face.

CCS CONCEPTS

• **Human-centered computing**;

KEYWORDS

interpersonal synchrony, wearable sensing, eye tracking

ACM Reference Format:
Aman Gupta, Finn L Strivens, Benjamin Tag, Kai Kunze, and Jamie A Ward. 2019. Blink as you Sync - Uncovering Eye and Nod Synchrony in Conversation using Wearable Sensing. In *Proceedings of PREPRINT: 2019 International Symposium on Wearable Computers (PREPRINT ISWC '19).* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

*First two authors contributed equally to the paper.

PREPRINT ISWC '19, September 9–13, 2019, London, United Kingdom

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

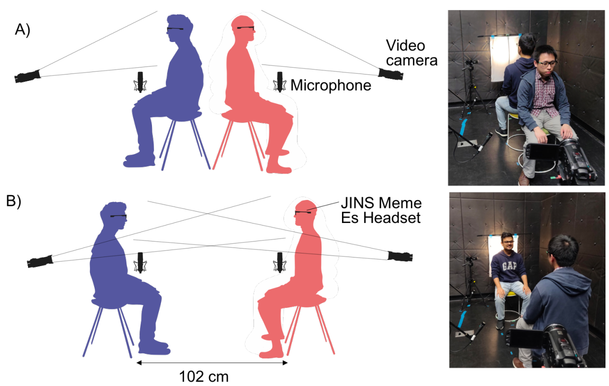

Figure 1: Experimental setup. Participants had one conversation back to back (A) and one conversation face to face (B). Audio and video recording directionally captured both conversations.

1 INTRODUCTION

Understanding a person's context and activity is at the core of making wearable computing more personal and transparent. Activity recognition covers a wide range of wearable modalities and applications, including physical (e.g., daily step counting, assembly task analysis), physiological (e.g., heart rhythm analysis, breathing rate), and cognitive (e.g., reading detection, cognitive load tracking). To-date, most of the focus of wearable sensing has been on the individual. However there is growing interest in the topic of social sensing, i.e. the in-situ capture and analysis of human social behavior using unobtrusive devices [20].

A particularly useful social signal is interpersonal synchrony. Interpersonal synchrony reveals the degree of temporal coordination between people during social interaction, and can provide cues on things like social engagement, and affect [2]. We propose measuring interpersonal synchrony using the head nods and eye blinks of dyads in conversation. Nods and blinks are used due to their importance as non-verbal signals [6], their connection to cognitive processes like sustained attention [16], and the relative ease of collecting this data using commercial head-worn wearables.

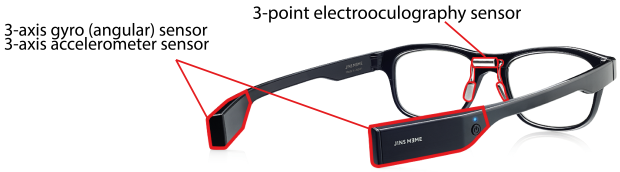

Figure 2: J!NS Meme glasses with 6-axis sensors and (EOG) electrodes.

We present a method of using smart eyewear to study non-verbal, interpersonal synchrony using eye and head movement during dyadic, open conversation. The work has two ultimate goals: to evaluate a new, wearable, approach for psychologists wishing to study interpersonal synchrony in-the-wild, and to provide a dataset of head and eye movement during conversation for researchers interested in building socially aware wearable applications [14]. The specific contributions include:

- A public dataset (http://eyewear.pro/blinksync) of 42 participants in conversation (10 in English, 10 in Japanese, 1 in Chinese), including head and eye movements via Inertial Measurement Unit (IMU) and Electrooculography (EOG), as well as synchronized audio and video.
- An application of wavelet coherence as a tool to analyse interpersonal sychrony on two different wearable sensing modalities (nods via IMU and blinks via EOG).
- Evaluation of two hypotheses: 1) we synchronise nods and blinks during conversation (true), and 2) we synchronise more face-to-face than back-to-back (false).

2 RELATED WORK

There is a long history in applied psychology dedicated to the study of nonverbal interaction using movement [9, 10, 27]. Tschacher et al. found a link between interpersonal synchrony and affect [27], revealing that if people move in sync, they tend to feel more positive towards one another. Ward et al. show interpersonal synchrony can be captured between actors and autistic children in theatre – and suggest how this might be used as a measure of social engagement [30].

Nodding is a particularly useful back-channel in conversation [11]. It is used to express attentiveness [17], as well as stance and affiliation [23]. Hale et al. use motion capture to reveal the importance of head nods as a mechanism for coordination during conversation [10]. They reveal a link between listening behaviour and fast nods of 1.5 to 5 Hz.

Eye blinks are a useful cognitive indicator [22, 25], but they also reveal information on social interaction. Nakano et al. found that when watching a video of a person speaking, listeners synchronise their eye blinks to those of the speaker, but only when the speaker's mouth and eyes are both visible [19]. They suggest a visual role of blinks and mouth movement in coordinating face-to-face conversation. This reinforces the early observations of Dittmann et al [4], which linked blink (and nod) synchrony to coordinating breakpoints in speech. The social importance of blinks was also demonstrated in a further work by Nakano et al. showing a lack of eye blink synchrony in autism [18], a condition characterised by communication difficulties.

Our work is closely related to the field of Social Sensing [20, 21]. This grew out of the initial efforts by Choudhury and Pentland to use wearable motion and audio to analyse group interactions [1]. Gordan et al. pursue collaborative activity recognition for group activities [7], while Ward et al. use audio and movement to uncover within-group collaborations [29]. In the field of tele-presence, Madan et al. show how the head nods of remote users can be reproduced in real time to facilitate group discussion [15].

Much of the work on non-verbal analysis of social signals use either computer vision, or pocket worn sensing (e.g. mobile phones) [12, 13, 28]. Wearables positioned on the body have the potential to provide a richer source of information on social signals.

To uncover synchrony between participants, we use a method based on wavelet cross-coherence. This method highlights correlation in both time and frequency between two signals, and was originally developed to measure co-variations in weather patterns [8]. It has since been used to measure interpersonal synchrony of head movements in conversation [5, 10]. And was applied to wrist-worn sensor data to detect social engagement [30]. The present work introduces a first attempt to apply wavelet coherence to conversation data obtained from two different head-mounted sensors (EOG and IMU).

3 EXPERIMENTAL SETUP

Pairs of participants were asked to have two conversations on two assigned topics. They had one conversation facing each other (FF) and another sat back to back (BB) so that they were unable to see the other person (Figure 1). Each conversation lasted exactly five minutes. The first conversation topic, adapted from Tschacher et al., was to "plan a 4-course meal together using only ingredients that neither of you like" [27]. The second topic was to "plan one day of a holiday only doing things that neither of you enjoy." The topics were described to participants directly before the beginning of each conversation. The direction faced for the first conversation was alternated throughout the study, and the two topics were used equally in both directions.

Apparatus and signals

All interactions were recorded using two video cameras and two channel audio recording (Figure 1). Each participant's EOG eye motion data, as well as IMU head movement data

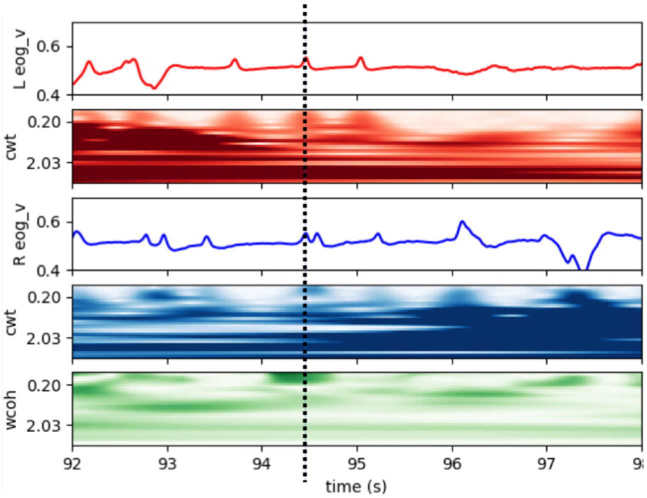

Figure 3: 6s example conversation (dyad 12, BB). Raw EOG-V signals shown, with corresponding continuous wavelet transforms (cwt) for each, and the resulting wavelet coherence WCOH spectrogram from combining these. Darker regions on the spectrograms show higher coherence values. Dotted line shows a moment of sychronous blinking and resulting wcoh (at scale of approx. 0.2s).

(accelerator and gyroscope), was recorded using J!NS MEME glasses. The glasses are equipped with EOG sensors around the nose, and 3-axis IMU in the frame. All logged data was transmitted to Android smartphones via Bluetooth LE (low energy). With a sampling rate of 100Hz, the glasses stream 10 datapoints (3 ACC, 3 GYRO, 4 EOG) at a time.

We use 'nod' as a shorthand for head pitch (y-axis acceleration, or ACC-Y) because nodding is the most common (and dominant) head pitch movement. Similarly we use blink as a shorthand for vertical EOG, or EOG-V. During a blink, the eyes perform a characteristic up- and downward motion that is expressed in the EOG-V signal [3]. This is created, in part, by changes in measuring the retinal-corneal potential, but also movement of the eyelid muscle. The characteristic voltage from blink is larger than any other eye movement, and is thus a dominant signal in EOG-V.

Applying cross-wavelet coherence directly to pairs of ACC-Y (and EOG-V) gives an indication that people are nodding (or blinking) in a coordinated way with one another: it highlights if the signals change at the same time.

Participants

We recorded 42 participants (22 female, average age 23, STD 4), who were predominantly university students. Participants registered themselves for the study in pairs or were assigned to dyads with someone fluent in their language to allow coherent communication. 21 pairs of participants (n=21 dyads) were asked to speak in their mutually preferred language. In total we had 10 English language pairs, 10 Japanese language pairs, and one Chinese language pair. All participants gave written consent after getting informed of the study design, setup, and potential data usage for analysis, including video and audio recordings before the experiment.

In the following analysis, we use data from 18 of the 21 dyads. Two excluded pairs were due to hardware synchronisation issues. The lone Chinese-language pair was also excluded (in part because one participant wore a face-mask, and we are interested in face-to-face effects).

4 WAVELET COHERENCE ANALYSIS

Wavelets allow us to decompose a signal into its frequency components while preserving temporal information, and without the need for windowing [26]. Obtaining the wavelet transform from two signals and then combining the outputs provides a way of obtaining the common time-spectral response. Two related methods of combining these include the cross-wavelet transform, which highlights the frequencies with high common power, and the wavelet coherence transform, which highlights common frequencies regardless of power [8]. Here we use wavelet coherence because of its superior performance on subtle, lower-power data.

The wavelet coherence spectrogram is obtained by combining wavelet spectrograms of the two signals being analysed (one from each of the conversing participants, here referred to as left, L and right, R). This process is shown using EOG-V for a 6s sequence of two people conversing in Figure 3. The wavelet coherence spectrogram (for both EOG-V and ACC-Y) is obtained in 3 steps: 1) low-pass filter the raw signals for L and R (5th-order Butterworth, cut-off 20Hz), 2) apply a continuous wavelet transform to the signals, W_L and W_R, 3) calculate the cross-wavelet transform by multiplying W_L by the complex conjugate of the other W_R^*, i.e. $W_{L,R} = W_L * W_R^*$, and then normalising for signal power to obtain the wavelet coherence (see [8] for full details). The wavelets used in this work are calculated using the continuous wavelet transform function, with a Morlet base, from the PyCWT module in Python (https://github.com/regeirk/pycwt).

Wavelet coherence spectrograms were computed for all the EOG and IMU data signals between conversing partners, and the results averaged for each condition over time to give a typical frequency response. Here we present only the data from vertical eye movement, EOG-V (blink), and y-direction acceleration, ACC-Y (head nods), as these are the relevant signals to the current study. The frequency response is represented by the approximate wavelet scale periods, or $1/frequency$. Paired t-tests (with p=0.05, N=18) were applied across 121 different wavelet scales. To account for multiple comparisons, we applied Benjamini-Hochberg, false discovery rate (FDR) correction at 0.05.

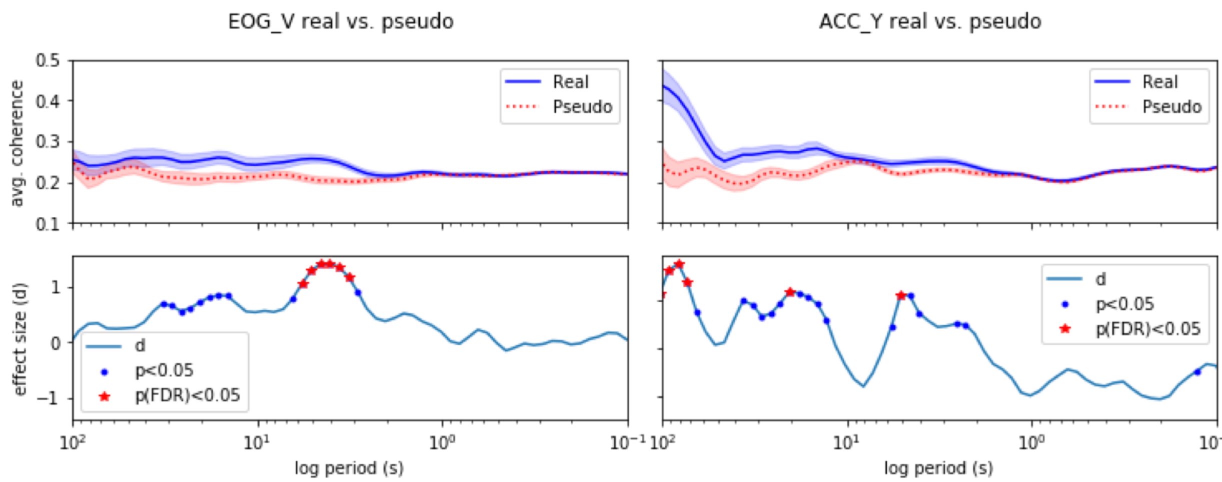

Figure 4: Real conversation vs. pseudo for blinks (EOG-V) and nods (ACC-Y). Average coherence for each condition is shown (with standard mean error, SME, in the shaded regions). Effect size (Cohen-d) is also shown with significance levels highlighted. This shows 1) we synchronise our blinks at periods of greater than 2s during conversation, and 2) we synchronise head nods over similar frequencies.

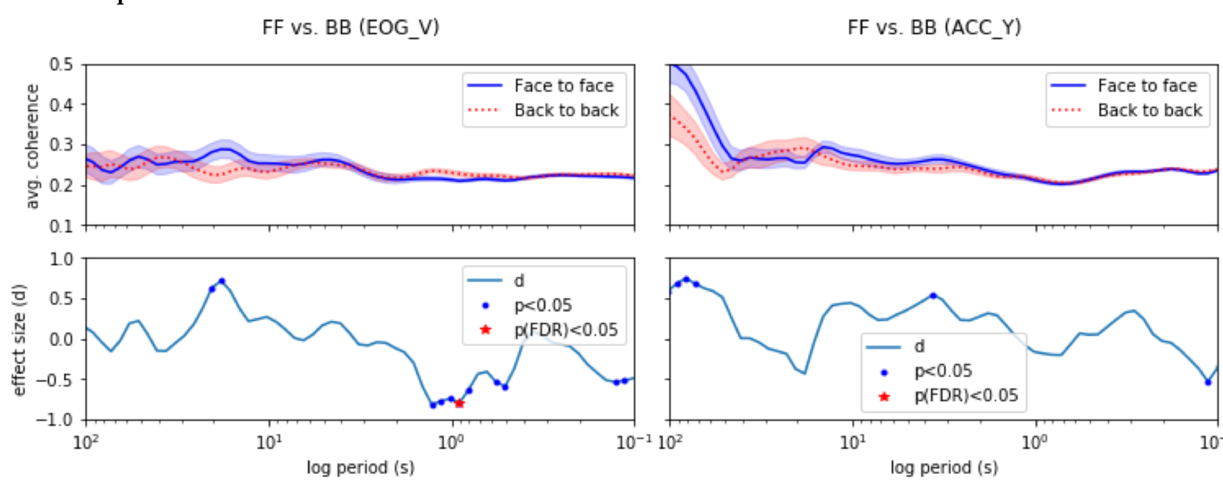

Figure 5: FF vs. BB conversation for EOG-V and ACC-Y. This shows 1) much of the synchrony in both nods and eye blinks occurs irrespective of whether people are FF or not, 2) and people synchronise their blinks at periods of 1 s more in BB.

5 RESULTS AND DISCUSSION

We performed evaluations to test 2 hypotheses: 1) people synchronise nods and blinks during conversation, 2) people synchronise more when they can see one another vs. when they cannot.

Do people synchronise in conversation?

Coherence data from real conversations were compared against coherence data from pseudo-conversations. Pseudo conversations approximate a random interaction by calculating the coherence of two signals taken from different conversations. To maintain the validity of the dependent t-tests, each pseudo conversation is calculated from the same participants used in the corresponding real conversation. To calculate a pseudo, we generate the coherence of person L (from FF), with their partner, R (from BB), and vice versa (L from BB, R from FF). We then average both to generate a single pseudo response. By comparing real-vs-pseudo in this way, we can uncover synchronicity that occurs in actual conversation as distinct from just the combination two individuals speaking.

Figure 4 shows our main result, which confirms the hypothesis that people synchronise with one another in eye blink, and to a slightly lesser extent in head nod. The effect is particularly strong for EOG-V at periods of 2 to 5 s, which roughly corresponds to typical eye blink rates in individuals.

Do people synchronise more face-to-face?

The right plots on Figure 5 suggest that there is generally no significant difference between conversants' ACC-V when face to face (FF) vs. back to back (BB).

However, when analysing EOG-V (the left plot), a significant difference is found in favour of BB at interaction periods of around 1s (1Hz). This surprising result suggests that people coordinate blinks more when they cannot see one another.

A rate of 1 Hz is faster than typical blink rates. It should be stressed that this analysis does not necessarily imply a constant stream of synchronised blinks every second. The result might also be explained by a series of occasional, synchronised, slow blinks (each lasting around 1s), or short bursts of 1 Hz blinks. Currently, we do not have a full explanation for this phenomenon, but hypothesize that either entrainment of breathing patterns of people who sit next to one another and talk, but cannot see one another, might be a factor. It may be also that the lack of visual connection leads to a higher dependence on auditory attention to be able to follow the conversation, which results in synchronized eye blinks due to increased engagement [19].

6 CONCLUSIONS AND FUTURE WORK

In this paper, we provide a data set exploring over 17 hours of natural conversations recorded by smart eyewear. We show – for the first time – that synchrony of physiological signals and non-verbal communication gestures can be detected using an unobtrusive, off-the-shelf wearable sensing device. We demonstrate how wavelet coherence analysis might be used to highlight coordination between wearable signals from interacting participants.

One limitiation of the work is our use of raw EOG-V and ACC-Y signals as a proxy for 'blink' and 'nod' behaviour. This is a valid assumption considering the dominance of nods and blinks on these signals. However, future work might consider a higher-level analysis, for instance using dynamic time warping on the output of separate blink and nod detectors.

In future work we intend to explore the cultural differences inherent in non-verbal behaviour. There are known differences in nodding between Japanese and English speaking populations, for example, which we will explore more fully in a follow-on work (e.g., [24]).

We also plan to extend the research beyond dyads to analyze group discussions. An important addition will be to test the robustness of these methods to in-the-wild situations and to evaluate real-time classification of conversational behaviour. This will allow eventual application of these methods to help build wearable aids that benefit from better awareness of human social context.

REFERENCES

[1] Tanzeem Choudhury and Alex Pentland. 2003. Sensing and modeling human networks using the sociometer. In *null*. IEEE, 216.

[2] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen. 2012. Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Transactions on Affective Computing* 3, 3 (July 2012), 349–365. https://doi.org/10.1109/T-AFFC.2012.12

[3] D Denney and C Denney. 1984. The eye blink electro-oculogram. *British Journal of Ophthalmology* 68, 4 (1984), 225–228. https://doi.org/10.1136/bjo.68.4.225

[4] Allen T Dittmann and Lynn G Llewellyn. 1968. Relationship between vocalizations and head nods as listener responses. *Journal of personality and social psychology* 9, 1 (1968), 79.

[5] Ken Fujiwara and Ikuo Daibo. 2016. Evaluating Interpersonal Synchrony: Wavelet Transform Toward an Unstructured Conversation. *Frontiers in psychology* 7 (2016).

[6] Marjorie Harness Goodwin and Charles Goodwin. 1986. Gesture and coparticipation in the activity of searching for a word. *Semiotica* 62, 1-2 (1986), 51âĂŞ75. https://doi.org/10.1515/semi.1986.62.1-2.51

[7] Dawud Gordon, Jan-Hendrik Hanne, Martin Berchtold, Ali Asghar Nazari Shirehjini, and Michael Beigl. 2013. Towards collaborative group activity recognition using mobile devices. *Mobile Networks and Applications* 18, 3 (2013), 326–340.

[8] Aslak Grinsted, John C Moore, and Svetlana Jevrejeva. 2004. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear processes in geophysics* 11, 5/6 (2004), 561–566.

[9] Uri Hadar, Timothy J Steiner, and F Clifford Rose. 1985. Head movement during listening turns in conversation. *Journal of Nonverbal Behavior* 9, 4 (1985), 214–228.

[10] Joanna Hale, Jamie A Ward, Francesco Buccheri, Dominic Oliver, and Antonia Hamilton. 2018. Are you on my wavelength? Interpersonal coordination in naturalistic conversations. (2018).

[11] Marie Helweg-Larsen, Stephanie J. Cunningham, Amanda Carrico, and Alison M. Pergram. 2004. To Nod or Not to Nod: An Observational Study of Nonverbal Communication and Status in Female and Male College Students. *Psychology of Women Quarterly* 28, 4 (2004), 358–361. https://doi.org/10.1111/j.1471-6402.2004.00152.x arXiv:https://doi.org/10.1111/j.1471-6402.2004.00152.x

[12] Dinesh Babu Jayagopi, Taemie Kim, Alex Sandy Pentland, and Daniel Gatica-Perez. 2010. Recognizing conversational context in group interaction using privacy-sensitive mobile sensors. In *Proceedings of the 9th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 8.

[13] Youngki Lee, Chulhong Min, Chanyou Hwang, Jaeung Lee, Inseok Hwang, Younghyun Ju, Chungkuk Yoo, Miri Moon, Uichin Lee, and Junehwa Song. 2013. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 375–388.

[14] P. Lukowicz, S. Pentland, and A. Ferscha. 2012. From Context Awareness to Socially Aware Computing. *IEEE Pervasive Computing* 11, 1 (January 2012), 32–41. https://doi.org/10.1109/MPRV.2011.82

[15] Anmol Madan, Ron Caneel, and Alex Sandy Pentland. 2004. GroupMedia: distributed multi-modal interfaces. In *Proceedings of the 6th international conference on Multimodal interfaces*. ACM, 309–316.

[16] Sander Martens and Brad Wyble. 2010. The attentional blink: Past, present, and future of a blind spot in perceptual awareness. *Neuroscience and Biobehavioral Reviews* 34, 6 (2010), 947–957. https://doi.org/10.1016/j.neubiorev.2009.12.005

[17] Anthony Mulac, Karen T. Erlandson, W. Jeffrey Farrar, Jennifer S. Hallett, Jennifer L. Molloy, and Margaret E. Prescott. 1998. âĂIJUh-huh. What's That All About?âĂİ: Differing Interpretations of Conversational Backchannels and Questions as Sources of Miscommunication Across Gender Boundaries. *Communication Research* 25, 6 (1998), 641–668. https://doi.org/10.1177/009365098025006004 arXiv:https://doi.org/10.1177/009365098025006004

[18] Tamami Nakano, Nobumasa Kato, and Shigeru Kitazawa. 2011. Lack of eyeblink entrainments in autism spectrum disorders. *Neuropsychologia* 49, 9 (2011), 2784–2790.

[19] Tamami Nakano, Yoshiharu Yamamoto, Keiichi Kitajo, Toshimitsu Takahashi, and Shigeru Kitazawa. 2009. Synchronization of spontaneous eyeblinks while viewing video stories. *Proceedings. Biological sciences / The Royal Society* 276, 1673 (2009), 3635–44. https://doi.org/10.1098/rspb.2009.0828

[20] Marianne Schmid Mast, Daniel Gatica-Perez, Denise Frauendorfer, Laurent Nguyen, and Tanzeem Choudhury. 2015. Social sensing for psychology: Automated interpersonal behavior assessment. *Current Directions in Psychological Science* 24, 2 (2015), 154–160.

[21] Erez Shmueli, Vivek K Singh, Bruno Lepri, and Alex Pentland. 2014. Sensing, understanding, and shaping social behavior. *IEEE Transactions on Computational Social Systems* 1, 1 (2014), 22–34.

[22] John A Stern, Donna Boyer, and David Schroeder. 1994. Blink Rate: A Possible Measure of Fatigue. *Human Factors* 36, 2 (1994), 285–297. https://doi.org/10.1177/001872089403600209

[23] Tanya Stivers. 2008. Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation. *Research on Language and Social Interaction* 41, 1 (2008), 31–57. https://doi.org/10.1080/08351810701691123 arXiv:https://doi.org/10.1080/08351810701691123

[24] Polly Szatrowski. 2000. Relation between gaze, head nodding and aizuti âĂŸback channelâĂŹat a Japanese company meeting. In *Annual Meeting of the Berkeley Linguistics Society*, Vol. 26. 283–294.

[25] Benjamin Tag, Andrew W Vargo, Aman Gupta, George Chernyshov, Kai Kunze, and Tilman Dingler. 2019. Continuous Alertness Assessments: Using EOG Glasses to Unobtrusively Monitor Fatigue Levels In-The-Wild. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 464:1—-464:12. https://doi.org/10.1145/3290605.3300694

[26] Christopher Torrence and Gilbert P Compo. 1998. A practical guide to wavelet analysis. *Bulletin of the American Meteorological society* 79, 1 (1998), 61–78.

[27] Wolfgang Tschacher, Georg M Rees, and Fabian Ramseyer. 2014. Nonverbal synchrony and affect in dyadic interactions. *Frontiers in psychology* 5 (2014), 1323.

[28] Kota Tsubouchi, Osamu Saisho, Junichi Sato, Seira Araki, and Masamichi Shimosaka. 2015. Fine-grained social relationship extraction from real activity data under coarse supervision. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers*. ACM, 183–187.

[29] J. A. Ward, G. Pirkl, P. Hevesi, and P. Lukowicz. 2017. Detecting physical collaborations in a group task using body-worn microphones and accelerometers. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. 268–273. https://doi.org/10.1109/PERCOMW.2017.7917570

[30] Jamie A Ward, Daniel Richardson, Guido Orgs, Kelly Hunter, and Antonia Hamilton. 2018. Sensing interpersonal synchrony between actors and autistic children in theatre using wrist-worn accelerometers. In *Proceedings of the 2018 ACM International Symposium on Wearable Computers*. ACM, 148–155.