Nonverbal Social Sensing: What Social Sensing Can and Cannot Do for the Study of Nonverbal Behavior From Video

Laetitia Aurelie Renier, Marianne Schmid Mast, Nele Dael, Emmanuelle Patricia Kleinlogel

frontiers in Psychology

REVIEW
published: 27 July 2021
doi: 10.3389/fpsyg.2021.606548

Nonverbal Social Sensing: What Social Sensing Can and Cannot Do for the Study of Nonverbal Behavior From Video

Laetitia Aurelie Renier, Marianne Schmid Mast, Nele Dael and Emmanuelle Patricia Kleinlogel*

Department of Organizational Behavior – Faculty of Business and Economics (HEC), University of Lausanne, Lausanne, Switzerland

The study of nonverbal behavior (NVB), and in particular kinesics (i.e., face and body motions), is typically seen as cost-intensive. However, the development of new technologies (e.g., ubiquitous sensing, computer vision, and algorithms) and approaches to study social behavior [i.e., social signal processing (SSP)] makes it possible to train algorithms to automatically code NVB, from action/motion units to inferences. Nonverbal social sensing refers to the use of these technologies and approaches for the study of kinesics based on video recordings. Nonverbal social sensing appears as an inspiring and encouraging approach to study NVB at reduced costs, making it a more attractive research field. However, does this promise hold? After presenting what nonverbal social sensing is and can do, we discussed the key challenges that researchers face when using nonverbal social sensing on video data. Although nonverbal social sensing is a promising tool, researchers need to be aware of the fact that algorithms might be as biased as humans when extracting NVB or that the automated NVB coding might remain context-dependent. We provided study examples to discuss these challenges and point to potential solutions.

Keywords: nonverbal behavior, social sensing, coding, extraction, communication, technology, annotations, algorithm

OPEN ACCESS

Edited by:
Judee K. Burgoon, University of Arizona, United States

Reviewed by:
Ross W. Buck, University of Connecticut, United States
Sophie Van Der Zee, Erasmus University Rotterdam, Netherlands

****Correspondence:***
Laetitia Aurelie Renier
laetitia.renier@unil.ch

Specialty section:
This article was submitted to Personality and Social Psychology, a section of the journal Frontiers in Psychology

Received: *15 September 2020*
Accepted: *21 June 2021*
Published: *27 July 2021*

Citation:
Renier LA, Schmid Mast M, Dael N and Kleinlogel EP (2021) Nonverbal Social Sensing: What Social Sensing Can and Cannot Do for the Study of Nonverbal Behavior From Video. Front. Psychol. 12:606548. doi: 10.3389/fpsyg.2021.606548

INTRODUCTION

Investigating nonverbal behavior (NVB), and in particular kinesics, namely face and body motions used in communication (Birdwhistell, 1955; Burgoon and Dunbar, 2018), involves observing social interactions and coding movements of participants in the face and the body. Manually coding NVB takes a considerable amount of time and resources because it means having coders sit in front of a video screen and, for instance, count the frequency of smiles, calculate the duration of gazing, code interruptions, or rate the target on a more global judgment (e.g., how dominant or deceiving) for many hours over many days. Moreover, this does not include the additional work of training the coders and establishing reliability among them.

Due to advanced growth in computer vision, new technologies and approaches (e.g., SSP, Vinciarelli et al., 2009a,b, 2012) have been developed to use and train algorithms to code NVB as action/motion units or as more global judgments (inferences) from videotaped individuals in

social interactions (e.g., trustfulness). This has given rise to nonverbal social sensing, an approach that allows to automatize most of the NVB coding.

Once such algorithms are developed, they have the advantage of being scalable. Therefore, to the extent that researchers code the same NVB or judge the same inferences in different studies, such algorithms are valuable to researchers. Moreover, there is no standardized codebook detailing exactly how to code NVB (e.g., should smiling be assessed as a frequency, a duration, or a general impression about how much a person smiles on a scale of 1–5), which makes the comparison of results pertaining to NVB difficult across different studies. If more researchers used nonverbal social sensing, this field might gain in standardization and we might discover new insights that were not previously possible since the different coding methods would introduce too much noise to detect the signal. Furthermore, using nonverbal social sensing, when studying NVB, has the potential to reveal meaningful nonverbal patterns more easily (e.g., looking at the interaction partner while speaking, see Burgoon et al., 2014 for an example in detection of deception using computer-assisted coding and an algorithm to identify temporal patterns) instead of extracting only isolated NVB cues (e.g., duration of looking at the interaction partner and the number of speech turns of the target). These advantages might attract new researchers to study NVB, thus enriching and broadening the field.

The aim of this paper is to provide information and guidance to researchers who consider using nonverbal social sensing for their studies. We explained how nonverbal social sensing works, where we see the challenges of using it for the study, and how we recommend addressing such challenges. We illustrated these aspects with selective study examples.

In this paper, we focused on kinesics and the use of nonverbal social sensing based on video recordings (see Poppe, 2017 for an application of nonverbal social sensing beyond video recordings). Kinesics refers to two categories of NVB: (1) gesture and posture and (2) face and eye behavior (Vinciarelli et al., 2009a; the latter is also referred to as gaze, Harrigan, 2005, p. 137). Moreover, we focused on the extraction of NVB or inferences based on videotaped targets. We did not consider the sensor-based technologies, which require participants to wear sensors that register their NVB during the interaction task (see, e.g., Poppe et al., 2014; Rahman et al., 2019).

THE LEVEL OF NONVERBAL CODING: UNIT VS. INFERENCE

We studied the NVB coding on two different levels: action/motion units and kineme/inferences. An action/motion unit refers to specific body motions, such as muscle movements in the face and frequency or duration of a specific NVB (e.g., head motion and movement of the lips) or in the body (e.g., arm movement and leaning). As for "micro-kinesics," these units do not carry social meaning (see Birdwhistell, 1952). However, researchers are interested not only in specific nonverbal cues but also in inferences and the coding of global judgments based on NVB. Coders make inferences about trustworthiness, hireability, charisma, personality, or motivation of a target by observing the behaviors of the participant.

The lower the level of abstraction in coding, the more the interpretation of what the behavior means is already included in the coding, whereas higher levels of abstraction need interpretation and information about the context (see Birdwhistell, 1970). To illustrate, the number of smiles does not have much meaning attached to it. The meaning of smiling depends largely on the context. For instance, the simulation-of-smiles model (Niedenthal et al., 2010; Rychlowska et al., 2017) proposes to distinguish smiles according to their roles as follows: the smile that communicates positive emotions (enjoyment smile), the smile that suggests positive social intentions (affiliative smile), and the smile that reflects status or control (dominance smile). However, coding friendliness for instance (which might be based on smiling, but not exclusively) involves coding the meaning of the underlying NVB (e.g., smile, eye contact, and voice tone) to decide to what extent an individual appears friendly.

In summary, action/motion units can be coded relatively objectively, whereas inferences are more subjective because they need interpretation and are more context-dependent. This distinction between units and inferences, between objective and subjective measurements (Burgoon and Dunbar, 2018), is key in understanding the workings and challenges of nonverbal social sensing.

HOW NONVERBAL SOCIAL SENSING WORKS

Nonverbal social sensing originates in the field of SSP. SSP aims at automatically analyzing and synthesizing social signals (Vinciarelli et al., 2009b). SSP allows transforming raw input data (e.g., video recordings of people in social interactions) into social signals (i.e., units or inferences). Developing algorithms for nonverbal social sensing requires input data (i.e., videos of participants and ground truth). The videos refer to the material on which the algorithm is trained to extract and classify the NVB. The ground truth refers to the labels (e.g., manual coding or self-report) used as the standard of extraction or classification. The ground truth is either collected for the entire dataset or only on a subset (i.e., training set) of videos.

Ground truth data can be obtained in many different ways. For instance, satisfaction ratings of clients of a call center have been used as ground truth to train an algorithm to predict client satisfaction based on vocal cues of the call center employees (Zweig et al., 2006; Segura et al., 2016). When wanting to develop an algorithm that extracts personality, self-reports or other reports of personality can be used as the ground truth or expert assessments. When interested in developing algorithms that mimic human perception and judgment (e.g., perceived trustworthiness and hireability), we required human coders who are instructed and trained to perform the coding manually (i.e., manual annotations serve as the ground truth) or naïve raters who report their perception of the targets (e.g., source credibility ratings, Pentland, 2018).

We present below the general functioning of nonverbal social sensing in the following sections. We first present the application to NVB studies at the unit level. Second, we present two approaches to address NVB at the inference level.

Nonverbal Social Sensing at the Unit Level

At the action/motion unit level, nonverbal social sensing allows capturing a wide variety of nonverbal cues, such as micro-expressions, gestures, and movements. To illustrate, in the case of micro-expressions, the coding consists of extracting the frequency and the duration of muscle movements in the face, such as in the study of facial expressions. One of the most well-known and used classification methods to manually code facial expressions is the facial action coding system (FACS; Ekman and Friesen, 1978). When using the FACS, human coders note whether a facial action (i.e., activation of facial muscles such as lip corners going up or brow-raising) is present when coding a video. From this coding system, researchers develop algorithms to automatically recognize facial action units (AUs) from still records (Pantic and Rothkrantz, 2004) and moving records (Kapoor et al., 2003; Bartlett et al., 2006; Tong et al., 2006). As an application example, researchers used nonverbal social sensing to study the existence of cross-cultural differences in smiling (AU12) and brow furrowing (AU4) (McDuff et al., 2017). These researchers used automated extraction of these two units to study the effect of culture (i.e., individualist vs. collectivist), setting (i.e., home vs. lab), and gender on facial expressions. Their use of nonverbal social sensing enabled them to observe cultural (e.g., higher rate of brow furrowing in individualist culture than in collectivist culture) and gender differences (e.g., more smiling and less brow furrowing for women than men in both cultures, but more pronounced differences in individualist culture) at a lesser cost and on a larger scale (e.g., using a sample of 740,984 participants across 12 countries). Some of these researchers particularly worked on the development of algorithms for the detection of AU12 and AU4 and on a corpus of data for the study of spontaneous facial expressions (McDuff et al., 2013).

We might also need human coders at the unit level. In order to train an algorithm to extract the number of times a person nods in a video, we need to define which head movements qualify as a nod. This information is typically provided by human coders. We need several independent human coders to watch the same videos and to judge whether a given head movement is a nod, and then, we need to test for reliability (i.e., the extent to which the independent coders are consistent). The machine is then fed with this information together with the corresponding video, and from these two inputs, the machine can learn to detect head nods (e.g., Nguyen et al., 2012). Once trained, the algorithm will have learned to extract the features and classify them as action/motion units and can be used on new datasets. However, instead of measuring the ground truth, researchers might also rely on open-source tools such as OpenPose (i.e., body behavior; Cao et al., 2019) or OpenFace (i.e., face behavior; Baltrusaitis et al., 2018). OpenPose is an open-source library for multi-person detection providing real-time pose estimation (e.g., head, hand, foot, and face). OpenFace is also an open-source library designed to detect facial landmarks (e.g., facial AU, head pose, and eye-gaze). Both libraries are well-recognized tools for coding NVB as action/motion units enabling researchers to skip the training stage of nonverbal social sensing (for an application of OpenFace, see Burgoon et al., 2021).

Nonverbal Social Sensing at the Inference Level

At the inference level, NVB is coded according to its meaning, starting from the kineme to a higher-order inference. As examples of kinemes, we cite visual dominance—the ratio of the percentage of looking while speaking divided by the percentage of looking while listening (Dovidio et al., 1988)—or visual back-channeling—head nods while listening (Nguyen et al., 2012).

Nonverbal social sensing allows extracting data related to higher-order inferences or global judgments. For example, algorithms can capture how dominant or how trustworthy individuals are perceived through the measure of a combination of NVB (Burgoon and Buller, 1994; Hall et al., 2005; Mast et al., 2011). For instance, researchers used nonverbal social sensing to automatically predict the level of dominance of individuals during group interactions (Jayagopi et al., 2009) or their hireability (Naim et al., 2015). Other instances include the detection of personality traits (e.g., Pianesi et al., 2008; Batrinca et al., 2011), using personality recognition to improve automated detection of deception (An et al., 2018), or the detection of emotions based on body movements (Glowinski et al., 2008).

For higher-order inferences, the following two main approaches are currently pursued. In the first approach, the NVB is extracted automatically from the video input (as described for the motion unit extraction), and this extracted NVB is then linked with the ground truth. The machine is trained to first extract the nonverbal features (e.g., a nod and a smile) and only then learns to link those to the higher-order inferences (e.g., the classification of a target as friendly). For instance, to predict who gets hired for a job, the machine can first extract a set of specific NVB and then link it to the ground truth of hiring decisions. Another example is training a machine to predict social skills or personality (Biel et al., 2013; Muralidhar et al., 2018; Rasipuram and Jayagopi, 2018) or emotions (Ahn et al., 2010) based on previously extracted nonverbal cues. Again, the ground truth has to be measured (e.g., human coders assessing the personality of the people in the video or a self-report of their personality). The machine that extracted the NVB will link the extracted NVB to the ground truth. This approach allows identifying the NVB that is conductive of being hired (Frauendorfer et al., 2014; Nguyen et al., 2014; Muralidhar et al., 2016), which is important for training and the transparency of the decision-making. When predicting that a person is conscientious, this approach allows knowing which NVB pattern is responsible for this prediction.

In the second approach, the machine is fed with the video input and the ground truth (e.g., hireability) and learns to classify the videos into (not) hireable without involving the explicit extraction of NVB. This second approach relies on deep learning (see Mehta et al., 2019 for a review of the use of deep learning

in the detection of personality traits). The machine is given the videos and the ground truth, which this time is an inference such as, for instance, how dominant a person behaves in a social interaction rated by external observers or the personality assessed *via* self-report. The machine learns the link between the training videos and the ground truth (i.e., annotated dataset). However, the researcher or user will not know which array of nonverbal cues the algorithm uses for the prediction. Does the machine judge people as dominant because they speak a lot, because of a loud tone of voice, because they move more, or because of their gender or skin color or any combination thereof? There is no way to be certain.

Using nonverbal social sensing for higher-order inferences by either first extracting the NVB or directly linking the videos to the ground truth (i.e., annotated dataset at the inference level) is a choice a researcher needs to make based on how important it is to know which behaviors are responsible for the inference. This approach might be considered less costly because researchers only need to feed the data to the machines without relying on human coders. However, the size of the dataset to be fed into the machine is large (i.e., hundreds of videos) and thus also potentially costly. Thus, the benefits and shortfalls of deep learning depend on the goals of the researchers. If they are interested in determining the behaviors responsible for the inferences, we cautioned researchers when using deep and unsupervised learning approaches given their black-box nature. However, if researchers are primarily interested in higher-order inferences, deep learning appears to be a suitable approach (e.g., Mehta et al., 2019). In between, supervised deep learning might also reduce the black-box aspect associated with unsupervised learning and might lead researchers to discover new patterns of behaviors and inferences (see LeCun et al., 2015). Finally, concerning lower-order inferences, advances in deep learning enable researchers to automatically extract human pose at a lesser cost (Mathis et al., 2018; Arac et al., 2019).

There are some corpora of annotated data concerning higher-order inferences available. For example, corpora of annotated data are available in the domain of group interaction studies (see Gatica-Perez, 2015 for a list of corpus), leadership emergence (corpus cited in Sanchez-Cortes et al., 2011, 2013), psychological distress (Gratch et al., 2014), or personality detection (Mana et al., 2007). These corpora might help reduce the cost of collecting the input data.

CHALLENGES WHEN USING NONVERBAL SOCIAL SENSING

Under this section, we highlight key challenges associated with the use of nonverbal social sensing for researchers. We additionally make suggestions to address them.

The Risk of Bias

Algorithms are often used because people think they are less biased. It is true that once the algorithm runs, it does not make a difference between, for example, women or men showing a certain behavior. It simply codes the behavior, whereas human coders might be affected by the gender of the person showing the behavior they are about to code. However, algorithms are only as good as the ground truth on which they are trained. In other words, if the ground truth is biased, the algorithm will be biased. The risk for biased ground truth is higher for predictions at the inference level than at the unit level because the former is a more subjective coding than the latter. Therefore, collecting ground truth on nodding is probably less biased than collecting ground truth on, for example, the hireability of a person for a job.

Bias might also plague algorithms that learn to detect patterns by themselves (i.e., unsupervised learning). For instance, algorithms might learn by themselves to discriminate women during the recruitment process (e.g., Dastin, 2018; Lambrecht and Tucker, 2019) without the developers or users being aware of this bias. To illustrate, an algorithm trained to select the best candidates for a job taught itself (i.e., based on the data fed to the algorithm) to discriminate against women during the recruitment process (Dastin, 2018). The algorithm extracted a rule based on the data it was fed (e.g., it detected a connection made between best candidates and males) and used the rule to make future judgments. This led Amazon to stop using its automated recruitment system. In the same vein, algorithms developed to attract new talents for STEM job opportunities targeted more men than women (Lambrecht and Tucker, 2019). As pointed out by Kleinberg et al. (2018), the training data might be "rooted in past discrimination" (Kleinberg et al., 2018, p. 116). Since the input data were biased, the output data were also biased.

Therefore, before using any established algorithms, researchers need to know what data the algorithm has been trained on to tentatively estimate the risk of bias. For example, if an algorithm has been trained to predict friendliness on videos showing mainly males from an individualistic culture, it is possible that the developed system will not offer accurate predictions for women or individuals from a collectivistic culture. In the same vein, researchers showed that algorithms trained on videos featuring only adults were biased in performing emotion recognition on a younger population (Howard et al., 2017). Researchers interested in developing their own algorithms also need to be critical about the input and output data used and created by their nonverbal social sensing system.

Biased decisions have important ethical ramifications. First, in the examples related to biased recruitment, the decision was made by a machine and not a human (see recommendations for trustworthy algorithms, High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019). Second, the algorithm ended up taking into account a feature protected by law (e.g., gender and ethnicity) to produce a decision that disadvantages the said group. Given that this subject is not the main focus of this study, we referred the reader to Kleinberg et al. (2018) for a discussion on the legal and ethical aspects of discrimination associated with the use of algorithms in the recruitment process and to Raghavan et al. (2020) for potential solutions and challenges.

Data Privacy

Another ethical issue is linked to data privacy. Social and computer scientists might not share the same ethical guidelines

when studying NVB. This difference might be aggravated by open-science policies. For instance, social scientists, studying NVB based on video recordings of participants, need to ensure the anonymity of the participants and to disclose the specific use of the collected data. Meanwhile, computer scientists might not be required to do the same and to obtain the consent of participants to reuse their data. In this context, sharing data or developing corpora useful for future studies might be more difficult to achieve for social scientists than for computer scientists. Still, following the Facebook-Cambridge Analytica scandal, an ethical crisis related to data protection has also shaken computer scientists. In this context, researchers need to be attentive to ethical compliance across fields of research. In this vein, fostering collaborations between social and computer scientists might help in determining ethical guidelines that are common to both fields.

Concerning ethical algorithms, we suggested that social scientists, interested in the use of nonverbal social sensing systems, should be well-informed about policies related to artificial intelligence (AI). For instance, in Europe, a group of experts was commissioned to work on ethical guidelines for AI (Biel et al., 2013). The requirements for the so-called trustworthy AI are (1) human agency and oversight, (2) technical robustness and safety, (3) privacy data and governance, (4) transparency, (5) diversity, non-discrimination, and fairness, (6) environmental and societal wellbeing, and (7) accountability. As suggested by the High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019), these seven requirements should be addressed, and reflected upon, if adherence is not feasible.

Context-Dependency of Nonverbal Social Sensing

The quality of the output generated using nonverbal social sensing depends on the extent to which the data coded by the algorithm resemble the data on which the algorithm had been trained. To illustrate, if researchers use an algorithm that extracts head nods and this algorithm has been trained on videos featuring people sitting in front of a camera, but the video material for which the researchers want to use the algorithm shows people from the side, instead of a frontal view, involved in social interaction, it is likely that the algorithm will not perform that well.

For inferences, context-dependency is even more of an issue and the extent to which inferences are domain-specific or transversal is unclear. Will an algorithm trained to extract personality from videos of targets self-presenting during a job interview extract personality from videos of people self-presenting for a dating site with equal accuracy? Will an algorithm trained to extract trustworthiness from videos of targets giving a public speech perform equally well on videos of people answering job interview questions?

We suggested to scholars, who want to use nonverbal social sensing, to gather information about the W5 + (i.e., where, what, when, who, why, and how of the video input data the algorithm has been trained on, Vinciarelli et al., 2009b) and on potential moderators (i.e., culture, relationship, and gender, Burgoon and Dunbar, 2018). This information will enable the researcher to gauge whether the algorithm can be used for this study, as well as highlight boundary conditions or limitations of the developed algorithms for future applications.

Off-the-Shelf vs. Tailored Approaches

Some nonverbal social sensing systems are readily available (i.e., OpenPose and OpenFace to code NVB as a unit or systems such as FaceReader to code NVB in the face as more more global judgment). These systems are easy to use for people outside the field of computer science. We thus encouraged researchers interested in coding NVB as action/motion units to try well-known off-the-shelf open-source solutions (e.g., OpenPose and OpenFace). However, researchers need to keep in mind that off-the-shelf systems might not be suited for their specific study purposes. For example, a researcher might need data on the duration of an NVB while off-the-shelf systems provide data on its frequency.

Nonverbal social sensing systems to code NVB at the inference level are also available on the market (e.g., FaceReader or Affectiva for facial expression, and HireVue and Pymetrics for hireability). These commercial off-the-shelf systems come with a caveat. They typically do not provide information about the input data (i.e., videos and ground truth) on which the algorithms have been trained, making it impossible to gauge the reliability and the accuracy of the inferences for the dataset of the researcher. To illustrate, the HireVue algorithm automatically generates a score of hireability and a rank to help companies make their hiring decisions. With this type of off-the-shelf solution, several questions arise: Does the algorithm take into account the protected features? Is human agency respected? Is the process transparent enough? How is accuracy assessed? To assess the quality of the inferences obtained by the off-the-shelf solutions, the researchers have to manually code a portion of their data and compare it with the output of the algorithm to ensure that the algorithm performs at the expected level.

Hence, when using an off-the-shelf system to code NVB at the inference level, researchers need to have access to its input and output data. This is necessary to assess its reliability and algorithm performance. Researchers are also advised to verify that the system is compliant with the existing guidelines on the use of AI (see the recommendation of OECD of the Council on Artificial Intelligence—OECD AI Principles; High-Level Expert Group on Artificial Intelligence (AI HLEG), 2019).

An alternative to the off-the-shelf solution is to become savvy in machine learning or to collaborate with computer scientists to develop an algorithm for automatic coding of NVB. These multidisciplinary collaborations can benefit both social and computer scientists by fostering the development of SSP and nonverbal social signals. Benefits have already been highlighted in the domain of neurosciences (Sedda et al., 2012). Social scientists can benefit from the technical expertise of computer scientists. Computer scientists can benefit from the expertise of social scientists in NVB studies (e.g., knowledge about taxonomies and key variables to take into account). Developing an algorithm

to code for NVB is only a viable solution if the developed algorithm can be used for other research projects. This is because the generation of the input data (i.e., videos and ground truth) and the machine learning process are time and resource-intensive.

To help identify the best nonverbal social sensing approach, researchers need a clear research question. This will help them determine the type of data and method that is needed. We suggest two complementary reflections. First, the general approach to study NVB must be clarified and operationalized. In this domain, we suggest following the pragmatic guide developed by Blanch-Hartigan et al. (2018) to identify the input data and the data collection method. This step is crucial to identify whether nonverbal a social sensing system is appropriate for the research project. The questions to be answered are: Is computer vision sufficiently developed to extract the NVB? Does a model to predict global judgment already exist? and Is it necessary to create a new nonverbal social sensing system? Second, to refine choices about coding NVB decisions, we suggest that researchers clarify their coding approach (Burgoon and Dunbar, 2018). Determining NVB coding strategies directly affects nonverbal social sensing. For instance, researchers interested in kinesics at the dyadic level need at least two cameras to record each member of the dyad for data capture. Another example of a decision that needs to be taken (i.e., when, where, and by whom) concerns the granularity of the temporal dimension. To illustrate, OpenPose enables researchers to automatically code the NVB for each second of the interaction. Other issues that need to be addressed include whether an off-the-shelf solution is available to code the macro-behaviors and whether researchers are interested in objective or subjective measurements in coding NVB as a unit or an inference.

CONCLUSION

Nonverbal social sensing can extract NVB from videotaped social interactions or it can make inferences based on NVB in videotaped social interactions. Both of these outputs are highly relevant for researchers, and because such algorithms allow scalability, they might attract new researchers in the domain of NVB, contributing to the advancement of the field.

However, these new technologies are still in development. Moreover, they are not free of biases and their input and output data are highly context-dependent. At this stage, ubiquitous sensing and automated extraction only complement human coding and particular caution, and scrutiny about the quality of the algorithm, needs to be taken before one can use these sensing and extraction technologies.

Researchers assessing the usefulness of nonverbal social sensing for their study should ask themselves the following questions: Can I use an algorithm that is already developed or do I have to develop my own? If I have to develop my own, do I have the necessary competencies or the necessary collaboration partners with those competencies? When using an existing algorithm: (a) Is the video input data similar to the training dataset? (b) How is the ground truth obtained? and (c) Do I know on which NVB the inferences are based? To ensure the quality and accuracy of the coding done by the algorithm on the data gathered by the researchers, said researchers might want to consider manually coding a subset of the data and then compare the performance of the algorithm with the manual coding.

The more established and robust algorithms for NVB extraction become, the more attractive they are for researchers to use and the more they might advance the field of NVB studies. This is because using established and robust algorithm for the automatic coding of NVB will improve the comparability of NVB across studies and has the potential to attract more researchers into the field.

AUTHOR CONTRIBUTIONS

LR, MSM, ND, and EK conceived of the presented idea. LR and MSM refined the main ideas and proof outline. All authors contributed different parts of writing to the manuscript with LR being in charge of coordinating and integrating and writing the most extensive part of the manuscript. All authors discussed the initial content and LR, MSM, and EK contributed to the final manuscript.

FUNDING

SNF Sinergia, CRSII5-183564.

REFERENCES

Ahn, S. J., Bailenson, J., Fox, J., and Jabon, M. (2010). "Using automated facial expression analysis for emotion and behavior prediction," in *The Routledge Handbook of Emotions and Mass meDia*, eds K. Doveling, C. von Scheve, and E. Konijn (New York, NY: Routledge), 349–369.

An, G., Levitan, S. I., Hirschberg, J., and Levitan, R. (2018). "Deep personality recognition for deception detection," in *Proceedings Interspeech*, Hyderabad, 421–425. doi: 10.21437/Interspeech.2018-2269

Arac, A., Zhao, P., Dobkin, B. H., Carmichael, S. T., and Golshani, P. (2019). DeepBehavior: a deep learning toolbox for automated analysis of animal and human behavior imaging data. *Front. Syst. Neurosci.* 13:20. doi: 10.3389/fnsys.2019.00020

Baltrusaitis, T., Zadeh, A., Lim, Y. C., and Morency, L. P. (2018). "Openface 2.0: facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, 59–66. doi: 10.1109/FG.2018.00019

Bartlett, M. S., Littlewort, G. C., Frank, M. G., Lainscsek, C., Fasel, I. R., and Movellan, J. R. (2006). Automatic recognition of facial actions in spontaneous expressions. *J. Mult.* 1, 22–35. doi: 10.4304/jmm.1.6.22-35

Batrinca, L. M., Mana, N., Lepri, B., Pianesi, F., and Sebe, N. (2011). "Please, tell me about yourself: automatic personality assessment using short self-presentations," in *Proceedings of the 13th International Conference on Multimodal Interfaces*, New York, NY: ACM, 255–262. doi: 10.1145/2070481.2070528

Biel, J. I., Tsiminaki, V., Dines, J., and Gatica-Perez, D. (2013). "Hi YouTube! personality impressions and verbal content in social

video," in *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction*, New York, NY: ACM, 119–126. doi: 10.1145/2522848.2522877

Birdwhistell, R. L. (1952). *Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture*. Lexington, KY: University of Kentucky Press.

Birdwhistell, R. L. (1955). Background to kinesics. *Review Gen. Semant.* 13, 10–17.

Birdwhistell, R. L. (1970). *Kinesics and Context: Essays on Body Motion*. Philadelphia: University of Pennsylvania Press.

Blanch-Hartigan, D., Ruben, M. A., Hall, J. A., and Schmid Mast, M. (2018). Measuring nonverbal behavior in clinical interactions: a pragmatic guide. *Patient Educ. Couns.* 101, 2209–2218. doi: 10.1016/j.pec.2018.08.013

Burgoon, J., and Dunbar, N. (2018). "Coding nonverbal behavior," in *The Cambridge Handbook of Group Interaction Analysis (Cambridge Handbooks in psychology)*, eds E. Brauner, M. Boos, and M. Kolbe (Cambridge: Cambridge University Press), 104–120. doi: 10.1017/9781316286302.007

Burgoon, J. K., and Buller, D. B. (1994). Interpersonal deception: III. Effects of deceit on perceived communication and nonverbal behavior dynamics. *J. Nonverbal Behav.* 18, 155–184. doi: 10.1007/BF02170076

Burgoon, J. K., Proudfoot, J. G., Schuetzler, R., and Wilson, D. (2014). Patterns of nonverbal behavior associated with truth and deception: illustrations from three experiments. *J. Nonverbal Behav.* 38, 325–354. doi: 10.1007/s10919-014-0181-5

Burgoon, J. K., Wang, X., Chen, X., Pentland, S. J., and Dunbar, N. E. (2021). Nonverbal behaviors "speak" relational messages of dominance, trust, and composure. *Front. Psychol.* 12:624177. doi: 10.3389/fpsyg.2021.624177

Cao, Z., Hidalgo, G., Simon, T., Wei, S. E., and Sheikh, Y. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* 43, 172–186. doi: 10.1109/TPAMI.2019.2929257

Dastin, J. (2018). *Amazon Scraps Secret ai Recruiting Tool That Showed Bias Against Women*. London: Reuters.

Dovidio, J. F., Brown, C. E., Heltman, K., Ellyson, S. L., and Keating, C. F. (1988). Power displays between women and men in discussions of gender-linked tasks: a multichannel study. *J. Personal. Soc. Psychol.* 55, 580–587. doi: 10.1037/0022-3514.55.4.580

Ekman, P., and Friesen, W. (1978). *Facial Action Coding System*. Bel Air, CA: Consulting Psychologists Press.

Frauendorfer, D., Schmid Mast, M., Nguyen, L. S., and Gatica-Perez, D. (2014). Nonverbal social sensing in action: unobtrusive recording and extracting of nonverbal behavior in social interactions illustrated with a research example. *J. Nonverbal Behav.* 38, 231–245. doi: 10.1007/s10919-014-0173-5

Gatica-Perez, D. (2015). Signal processing in the workplace. *IEEE Signal Proc. Mag.* 32, 121–125. doi: 10.1109/MSP.2014.2359247

Glowinski, D., Camurri, A., Volpe, G., Dael, N., and Scherer, K. (2008). "Technique for automatic emotion recognition by body gesture analysis," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Anchorage, AK, 1–6. doi: 10.1109/CVPRW.2008.4563173

Gratch, J., Artstein, R., Lucas, G. M., Stratou, G., Scherer, S., Nazarian, A., et al. (2014). "The distress analysis interview corpus of human and computer interviews," in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, 3123–3128.

Hall, J. A., Coats, E. J., and LeBeau, L. S. (2005). Nonverbal behavior and the vertical dimension of social relations: a meta-analysis. *Psychol. Bull.* 131, 898–924. doi: 10.1037/0033-2909.131.6.898

Harrigan, J. A. (2005). "Proxemics, kinesics, and gaze," in *The New Handbook of Methods in Nonverbal Behavior Research*, eds J. A. Harrigan, R. Rosenthal, and K. Scherer (Oxford: Oxford University Press), 137–198.

High-Level Expert Group on Artificial Intelligence (AI HLEG) (2019). *Ethics Guidelines for Trustworthy AI*. European Commission Report. Avaliable online at: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (accessed March 15, 2021).

Howard, A., Zhang, C., and Horvitz, E. (2017). "Addressing bias in machine learning algorithms: a pilot study on emotion recognition for intelligent systems," in *Proceedings of IEEE Workshop on Advanced Robotics and Its Social Impacts (ARSO)*, Austin, TX, 1–7. doi: 10.1109/ARSO.2017.8025197

Jayagopi, D. B., Hung, H., Yeo, C., and Gatica-Perez, D. (2009). Modeling dominance in group conversations using nonverbal activity cues. *IEEE Trans. Audio Speech Lang. Proc.* 17, 501–513. doi: 10.1109/TASL.2008.2008238

Kapoor, A., Qi, Y., and Picard, R. W. (2003). "Fully automatic upper facial action recognition," in *2003 IEEE International SOI Conference*, Newport Beach, CA, 195–202.

Kleinberg, J., Ludwig, J., Mullainathan, S., and Sunstein, C. R. (2018). Discrimination in the age of algorithms. *J. Legal Anal.* 10, 113–174. doi: 10.1093/jla/laz001

Lambrecht, A., and Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Manag. Sci.* 65, 2966–2981. doi: 10.1287/mnsc.2018.3093

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Mana, N., Lepri, B., Chippendale, P., Cappelletti, A., Pianesi, F., Svaizer, P., et al. (2007). "Multimodal corpus of multi-party meetings for automatic social behavior analysis and personality traits detection," in *Proceedings of the 2007 Workshop on Tagging, Mining and Retrieval of Human Related Activity information*, (New York, NY: ACM), .9–14. doi: 10.1145/1330588.1330590

Mast, M. S., Hall, J. A., Cronauer, C. K., and Cousin, G. (2011). Perceived dominance in physicians: are female physicians under scrutiny? *Patient Educ. Couns.* 83, 174–179. doi: 10.1016/j.pec.2010.06.030

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., et al. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* 21, 1281–1289. doi: 10.1038/s41593-018-0209-y

McDuff, D., Girard, J. M., and el Kaliouby, R. (2017). Large-scale observational evidence of cross-cultural differences in facial behavior. *J. Nonverbal Behav.* 41, 1–19. doi: 10.1007/s10919-016-0244-x

McDuff, D., Kaliouby, R., Senechal, T., Amr, M., Cohn, J., and Picard, R. (2013). "Affectiva-mit facial expression dataset (am-fed): naturalistic and spontaneous facial expressions collected," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Piscataway, NJ: IEEE, 881–888.

Mehta, Y., Majumder, N., Gelbukh, A., and Cambria, E. (2019). Recent trends in deep learning based personality detection. *Artif. Intel. Rev.* 53, 2313–2339.

Muralidhar, S., Nguyen, L. S., Frauendorfer, D., Odobez, J. M., Schmid Mast, M., and Gatica-Perez, D. (2016). "Training on the job: behavioral analysis of job interviews in hospitality," in *ICMI 2016 - Proceedings of the 18th ACM International Conference on Multimodal Interaction*, (New York, NY: ACM), 84–91. doi: 10.1145/2993148.2993191

Muralidhar, S., Schmid Mast, M., and Gatica-Perez, D. (2018). A tale of two interactions: inferring performance in hospitality encounters from cross-situation social sensing. *Proc. ACM Interact. Mobile Wearable Ubiquitous Technol.* 2, 1–24. doi: 10.1145/3264939

Naim, I., Tanveer, M. I., Gildea, D., and Hoque, M. E. (2015). "Automated prediction and analysis of job interview performance: the role of what you say and how you say it," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Vol. 1, Ljubljana, 1–6. doi: 10.1109/FG.2015.7163127

Nguyen, L. S., Frauendorfer, D., Schmid Mast, M., and Gatica-Perez, D. (2014). Hire me: computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Trans. Multimed.* 16, 1018–1031. doi: 10.1109/TMM.2014.2307169

Nguyen, L. S., Odobez, J. M., and Gatica-Perez, D. (2012). "Using self-context for multimodal detection of head nods in face-to-face interactions," in *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, New York, NY: ACM, 289–292. doi: 10.1145/2388676.2388734

Niedenthal, P. M., Mermillod, M., Maringer, M., and Hess, U. (2010). The simulation of smiles (SIMS) model: embodied simulation and the meaning of facial expression. *Behav. Brain Sci.* 33, 417–433. doi: 10.1017/S0140525X10000865

Pantic, M., and Rothkrantz, L. J. M. (2004). Facial action recognition for facial expression analysis from static face images. *IEEE Trans. Syst. Man Cybernet. Part B Cybernet.* 34, 1449–1461. doi: 10.1109/TSMCB.2004.825931

Pentland, S. J. (2018). *Human-Analytics in Information Systems Research and Applications in Personnel Selection*. Doctoral dissertation. Tucson, AZ: The University of Arizona. UA Campus Repository.

Pianesi, F., Mana, N., Cappelletti, A., Lepri, B., and Zancanaro, M. (2008). "Multimodal recognition of personality traits in social interactions," in *Proceedings of the 10th International Conference on Multimodal Interfaces*, Chania, 53–60. doi: 10.1145/1452392.1452404

Poppe, R. (2017). "Automatic analysis of bodily social signals," in *Social Signal Processing*, eds J. Burgoon, N. Magnenat-Thalmann, M. Pantic, and A. Vinciarelli (Cambridge, MA: Cambridge University Press), 155–167. doi: 10.1017/9781316676202.012

Poppe, R., Van Der Zee, S., Heylen, D. K., and Taylor, P. J. (2014). AMAB: automated measurement and analysis of body motion. *Behav. Res. Methods* 46, 625–633. doi: 10.3758/s13428-013-0398-y

Raghavan, M., Barocas, S., Kleinberg, J., and Levy, K. (2020). "Mitigating bias in algorithmic hiring: evaluating claims and practices," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, New York, NY, 469–481. doi: 10.1145/3351095.3372828

Rahman, A., Clift, L. G., and Clark, A. F. (2019). "Comparing gestural interfaces using kinect and OpenPose [Poster presentation]," *in Proceedings of the 37th Computer Graphics & Visual Computing, Bangor, United Kingdom*, Bangor, 103–104. doi: 10.2312/cgvc.20191264

Rasipuram, S., and Jayagopi, D. B. (2018). Automatic assessment of communication skill in interview-based interactions. *Mult. Tools Appl.* 77, 18709–18739. doi: 10.1007/s11042-018-5654-9

Rychlowska, M., Jack, R. E., Garrod, O. G., Schyns, P. G., Martin, J. D., and Niedenthal, P. M. (2017). Functional smiles: tools for love, sympathy, and war. *Psychol. Sci.* 28, 1259–1270. doi: 10.1177/0956797617706082

Sanchez-Cortes, D., Aran, O., and Gatica-Perez, D. (2011). "An audio visual corpus for emergent leader analysis," in *Workshop on Multimodal Corpora for Machine Learning: Taking Stock and Road Mapping the Future, ICMI-MLMI*, Alicante.

Sanchez-Cortes, D., Aran, O., Jayagopi, D. B., Mast, M. S., and Gatica-Perez, D. (2013). Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *J. Mult. User Interf.* 7, 39–53. doi: 10.1007/s12193-012-0101-0

Sedda, A., Manfredi, V., Bottini, G., Cristani, M., and Murino, V. (2012). Automatic human interaction understanding: lessons from a multidisciplinary approach. *Front. Hum. Neurosci.* 6:57. doi: 10.3389/fnhum.2012.00057

Segura, C., Balcells, D., Umbert, M., Arias, J., and Luque, J. (2016). *"Automatic speech feature learning for continuous prediction of customer satisfaction in contact center phone calls," in Proceedings of the International Conference on Advances in Speech and Language Technologies for Iberian Languages, Lisbon, Portugal*, (New York, NY: Springer International Publishing), 255–265.

Tong, Y., Liao, W., and Ji, Q. (2006). "Inferring facial action units with causal relations," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, New York, NY, 1623–1629. doi: 10.1109/CVPR.2006.154

Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., et al. (2012). Bridging the gap between social animal and unsocial machine: a survey of social signal processing. *IEEE Trans. Affect. Comput.* 3, 69–87. doi: 10.1109/T-AFFC.2011.27

Vinciarelli, A., Pantic, M., and Bourlard, H. (2009a). Social signal processing: survey of an emerging domain. *Image Vis. Comput.* 27, 1743–1759. doi: 10.1016/j.imavis.2008.11.007

Vinciarelli, A., Salamin, H., and Pantic, M. (2009b). "Social signal processing: understanding social interactions through nonverbal behavior analysis," in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami, FL, 42–49. doi: 10.1109/CVPRW.2009.5204290

Zweig, G., Siohan, O., Saon, G., Ramabhadran, B., Povey, D., Mangu, L., et al. (2006). "Automated quality monitoring for call centers using speech and NLP technologies," *in Proceedings of the 2006 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology: Companion Volume: Demonstrations, New York City, USA,* New York, NY, 292-295.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.