The Role of Different Types of Conversations for Meeting Success

Ke Zhou and Marios Constantinides, *Nokia Bell Labs Cambridge, Cambridge, CB3 0FA, U.K.*
Luca Maria Aiello, *IT University of Copenhagen, 2300 Kùbenhavn, Denmark*
Sagar Joglekar and Daniele Quercia, *Nokia Bell Labs Cambridge, Cambridge, CB3 0FA, U.K.*

While current meeting tools are able to capture key analytics from both text and voice (e.g., meeting summarization), they do not often capture important types of conversations (e.g., a heated discussion resulting in a conflict being resolved). We developed a framework that not only analyzes text and voice, but also quantifies fundamental types of conversations. Upon analyzing 72 hours of conversations from 85 real-world virtual meetings together with their 256 self-reported meeting success scores, we found that our quantification of types of conversations (e.g., social support, conflict resolution) was more predictive of meeting success than traditional voice and text analytics. These new techniques will be essential to uncover patterns in online meetings that might otherwise go unnoticed.

Data analytics might help online participants run their meetings more efficiently through accurate and real-time feedback on logistics, attendees, and environment.[2,6] For example, a meeting tool could improve awareness of a meeting's atmosphere by visualizing participants' contributions and salient moments.[2] Current meeting tools already offer insights through audiovisual or textual support. For instance, an existing speech-based tool is able to mark important "action items" in the spoken dialogue,[15] while other tools are able to identify discussion and agreement within multiparty conversations.[10]

While such tools often rely on textual or audiovisual analyses, they do not capture all the aspects characterizing successful meetings. This is especially true for online meetings in which certain social cues (e.g., head movements[14] and body languages[4]) might go missing. Consider, for example, a virtual meeting during which only the host and a few participants enable the camera feed. In such a setting, while audio may convey, to a great extent, the sentiment and the prosody of the spoken words, the lack of physical presence and interaction makes it difficult to capture key conversations (e.g., a conversation expressing support); a situation that many might have experienced during the COVID-19 pandemic. As more meetings are held in virtual rooms whose content can be recorded, we are faced with an unprecedented opportunity to automatically analyze their characteristics, and understand their language's nuances. This work took that opportunity and, in so doing, made three main contributions:[a]

1) We collected 72 hours of meeting conversations from 85 real-world virtual meetings, held on Cisco's WebEx in a corporate setting. Additionally, we collected 256 self-reported meeting success scores, which we used as the ground truth in our predictive models (see the "Dataset" section).
2) Using our dataset, we developed metrics based on the literature (see the "Methodology" section), which capture traditional textual and verbal analytics, and newly defined types of conversations expressed in the spoken dialogue aiming at universally describing any type of social relationship. We built a model that predicts a meeting's success upon these metrics, and found that the quantification of different types of conversations was more predictive than

1536-1268 © 2021 IEEE
Digital Object Identifier 10.1109/MPRV.2021.3115879
Date of publication 10 November 2021; date of current version 3 December 2021.

[a]http://social-dynamics.net/FabConvo

traditional verbal and textual analytics (see the "Results" section).

3) We discuss potential uses of this new set of analytics in current and future tools for monitoring and improving the productivity in meetings (see the "Discussion and Conclusion" section).

RELATED WORK

Meeting analytics are key to a meeting's productivity;[2] in a sense, they provide a way to reflect and, ultimately, run meetings more effectively. They have been used for a variety of purposes, including offering post-meeting assistance,[15,20] providing real-time feedback to organizers or participants,[2,6,19] and quantifying human behavior to track whether a meeting was productive.[4,10,14]

Technologies for post-meeting assistance mainly focused on helping participants and organizers summarize past meetings by annotating meeting transcripts and detecting action items,[15] and by detecting sentiments.[20] The main objective was to allow participants to learn from past meetings. Other technologies focused on providing real-time feedback alongside a meeting. For example, Sarda *et al.*[19] developed a real-time feedback system highlighting speaker turns to foster more inclusive meetings. Finally, several studies focused on tracking behavior, including tracking the impact of micro conversational events (e.g., turn taking and transitions) on macro group performance,[10] and monitoring sensor outputs[4] that captured body postures and gestures that might impact a meeting's experience.

Most of the above studies leveraged a variety of signals, from visual (e.g., head movements[4,14]) to physical (e.g., heart rate[4]) to verbal (e.g., speech speed and linguistic properties[19]) to interactive (e.g., likes[2]). However, most of those previous studies fall short in understanding the relationship of those metrics with meeting experience; contrary to previous research, our study correlates meeting metrics with participants' self-reported experience. Additionally, most of prior work relies on audio transcripts, and often overlooks communication nuances and subtle cues.[9] That is why this work set out to find the relationship between social, verbal, sentiment cues, and meeting success.

DATASET

Using a Cisco's WebEx companion platform,[2] we collected data from 85 virtual corporate meetings with the consent of the participants. In total, these meetings lasted 4373 minutes with a median of 4 people participating in each meeting (min: 2, max: 65, with 11 meetings participated by more than 10 people). The dataset is comprised of a diverse range of meetings with varying duration (min: 20.6 minutes, median: 48.3 minutes, max: 180.2 minutes), hours of day (earliest and latest meeting happened at 8 A.M. and 6 P.M., respectively, on that day), days of week (Mon–Fri), and days of month (1–31). These meetings lasted for about 49 minutes on average, and all of them were conducted during business hours (8 A.M. to 6 P.M., Mon–Fri). The companion platform allowed people participants to earmark key moments with a mobile app. These moments were then converted into one minute long audio chunks, which the meeting participants could playback in retrospect to get a quick audio summary of the meeting. Earmark moments define salient moments, singling out important parts of the meetings.[2] The companion platform also allowed us to obtain self-reports. More specifically, at the end of each meeting, the participants were prompted to answer two questions: one captured $Q_{psychological}$, which is the extent to which *[a participant]* felt listened or motivated to be involved, and the other captured $Q_{execution}$, which is the extent to which *[a participant]* felt that the meeting had a clear purpose and structure. The two questions were answered on a 1–7 Likert-scale, with 7 indicating greater extent. These two questions resulted from an extensive large-scale crowdsourcing study that determined the key predictors of a meeting's psychological experience,[6] and are generalizable and independent from the specific analytics under study here.

Each meeting in the dataset was stored as a set of one-minute earmarked audio chunks, and of each participant's two self-reported answers. We transcribed the earmarked audio chunks using the state-of-the-art Google's API Speech-to-Text service;[b] each meeting's transcript was used in our textual analyses, while a meeting's audio was used in our audio analyses. In total, all the 85 meetings contained 1007 earmarked moments (a meeting on average had 11 earmarked moments), and 256 answers to the two questions.

METHODOLOGY

Using our collected dataset, we designed five metrics based on the literature that capture both verbal

[b]Speech-to-Text API: https://cloud.google.com/speech-to-text. It has been found that Google has superior performance on speech recognition compared to other platforms and tools.[13]

analytics (state-of-the-art) and types of conversations (our proposal). Verbal analytics metrics are denoted with (V). To allow for experimental comparison, we developed two additional state-of-the-art metrics based on textual analyses, which are denoted with (T).

(A) State-of-The-Art Meeting Analytics

1) *Content (T):* Following the work of Murray,[16] we considered a bag-of-words model that quantifies the frequencies of the most frequent uni-grams and bi-grams used in the meeting transcripts. To reduce sparsity, we counted the uni-grams and bi-grams that occur five times or more in the training set.
2) *Sentiment (T):* We applied sentiment analysis to capture the spectrum of sentiment expressed throughout the meeting. More specifically, we applied both VADER (rule-based)[12] and FLARE (based on deep-learning)[1] to the meeting transcripts.
3) *Sentiment (V):* Verbal sentiment has been linked to people's perception of a meeting's experience.[16] We used a deep-learning speech-based sentiment classifier[8] to extract verbal sentiment for each meeting. The classifier was trained on an audio dataset annotated with eight emotions: neutral, calm, happy, sad, angry, fearful, surprise, and disgust. The adopted classifier was shown to achieve empirically superior performance, obtaining a weighted average F1 score of 0.91 on a widely used public dataset.[8]
4) *Emotions From Pitch and Energy (V):* In verbal communication, pitch expresses emotional and paralinguistic information; it conveys emphasis, contrast, and intonation. Coutinho and Dibben[7] showed that prosodic features (e.g., pitch and energy) provide a reliable indication of the emotional status of conversational interactions. For example, the arousal state of a speaker (high activation versus low activation) affects the overall energy, and the energy distribution across the frequency spectrum.[7] To capture pitch and energy intensity patterns, for each meeting, we extracted the mean, the median, the standard deviation, the maximum, the minimum, and the range (max–min) of both the fundamental frequency and the energy. We also calculated the ratio of the up-slope of the pitch contour to that of the down-slope, which captures the fraction of high pitched voice regions.
5) *Emotions From Speech Rate (V):* The arousal state of a speaker has been found to affect the frequency and duration of pauses. For example, an unusually high speaking rate has been linked to altered emotional states.[17] To capture speech rate, we used: *i)* the number of syllables per duration, *ii)* the number of syllables per phonation time, and *iii)* the ratio of duration of voiced and unvoiced regions.
6) *Emotions From Prosody (V):* In addition to time-dependent acoustic features (e.g., pitch, energy, and speech rate), spectral features are often selected as a short-time representation for speech signal. It is known that, during meetings, happy utterances have higher energy at high frequency range, while sad utterances have lower energy at the same frequency range.[18] For each meeting, we computed the mel-frequency cepstrum (MFC) as it is a widely used representation of such short-term sound power spectrums.[18]

(B) Our Proposal: Types of Conversations

Choi *et al.*[5] and Deri *et al.*[9] showed that there are 10 dimensions that capture, to a great extent, the type of social interactions in a wide variety of communication types in the workplace (in, e.g., corporate email exchanges). These dimensions were found to universally describe any types of social relationships based on an extensive review of decades' worth of findings in sociology and social psychology.[9]

These dimensions[5] include: *knowledge* (exchange of ideas or information; learning, teaching), *power* (having power over the behavior and outcomes of another), *status* (conferring status, appreciation, gratitude, or admiration upon another), *trust* (will of relying on the actions or judgments of another), *support* (giving emotional or practical aid and companionship), *romance* (intimacy among people with a sentimental or sexual relationship), *similarity* (shared interests, motivations, or outlooks), *identity* (shared sense of belonging to the same community or group), *fun* (experiencing leisure, laughter, and joy), and *conflict* (diverging views, and conflict resolution).

Although these categories are not meant to cover exhaustively all possible social experiences, Deri *et al.*[9] provided empirical evidence that most people are able to characterize the nature of their relationships using these 10 concepts only. Through a crowdsourcing experiment, they asked people to spell out keywords that described their social connections, and found that all of them fitted into the 10

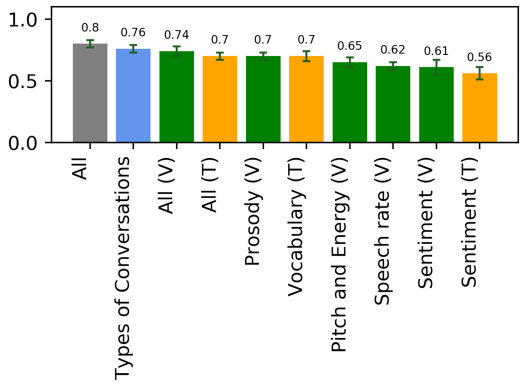

FIGURE 1. Evaluation (AUC) of our models trained on textual analytics (orange), verbal analytics (green), and types of conversations (blue). Model performances on the training dataset are reported on leave-one-out cross-validation folds, whereas error bars show the variance across all folds.

dimensions. We developed an long short-term memory (LSTM) classifier,[5] a type of recurrent neural network particularly suited to process data that is structured in temporal or logical sequences, to derive the 10 types of conversations. We selected the LSTM model as it has been shown to perform best compared to other state-of-the-art approaches.[5] The training data were acquired through crowdsourcing by labeling the 10 social dimensions on comments from a public discussion platform (Reddit) and from a corporate email exchange dataset (Enron), in total consisting of around 9k pieces of text. The 10 trained classifiers were shown to be general and robust, performing well on many different contexts, including social media, movie, corporate email exchanges and discussion forums.[5] We adopted such classifiers to quantify the types of conversations that typically occur in a meeting. We excluded the dimension of *romance* as it was not present in our meeting data.

Since our metrics might not be exhaustive to cover all types of meetings, we set out to test the extent to which these metrics are predictive of self-reported meeting success.

Self-Reported Meeting Success Score

To see how we defined a "success" score, consider a previous large-scale crowdsourcing study.[6] In it, a 28-item questionnaire was administered to 363 individuals whose answers were statistically analyzed through principal component analysis. The analysis showed that two factors were sufficient to mostly capture whether a meeting was successful or not: (a) the extent to which participants felt listened during the meeting or motivated to be involved ($Q_{\text{psychological}}$), and (b) the extent to which the meeting had a clear purpose and structure ($Q_{\text{execution}}$). To this end, we obtained the loading factors of these two questions, and used these loadings and the self-reports to compute an aggregated score for each attendee as: $\text{success} = (0.759 \cdot Q_{\text{psychological}}) + (0.673 \cdot Q_{\text{execution}})$. We binarized each meeting's success (using the median computed across all meetings, min: 5.5, median: 7.8, max: 10.0), and accordingly assigned the meeting to either a positive class or a negative one (i.e., categorizing all meetings to be "successful" or "unsuccessful").

RESULTS

To test the predictive power of our metrics, we developed classifiers for meeting success. We deployed a logistic regression, a support vector machine, a random forest, a XGBoost, and an AdaBoost classifier. We chose these classifiers as they represent a wide range of linear and nonlinear classification algorithms. These algorithms are also proven to be robust and perform well across datasets and applications. Based on our analyses, we found that the best performing model was AdaBoost, which is an ensemble learning method (also known as "meta-learning"). AdaBoost uses an iterative approach to learn from the misclassifications of weak classifiers, and builds a strong classifier by combining multiple weak classifiers; for brevity, we report only its results. We measured performance using a standard classification metric, that is, the area under curve (AUC), and employed a leave-one-out cross-validation.[c] We report the averaged AUC and the variance across all folds. To compare different models, we performed pair-wise t-test to determine to which extent their AUC values were statistically different from each other. *General Evaluation*. Figure 1 reports the AUC values for our models trained on different combinations of our metrics.

› By inspecting each individual metric independently, we found that "types of conversations" achieved the highest AUC of 76%, whereas the textual sentiment (T) metric, yielded the lowest AUC score of 56%. This is largely because most of the meeting transcripts do not contain explicit expressions of emotions. The "types of conversations" metric was found to statistically

[c]Leave-one-out cross-validation results in a more robust estimate of model performance on a small dataset.

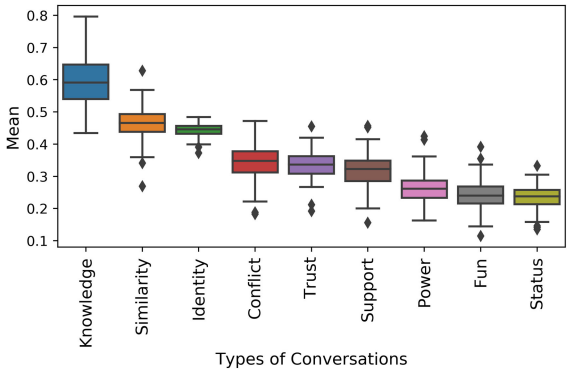

FIGURE 2. Distributions of types of conversations expressed in our meetings.

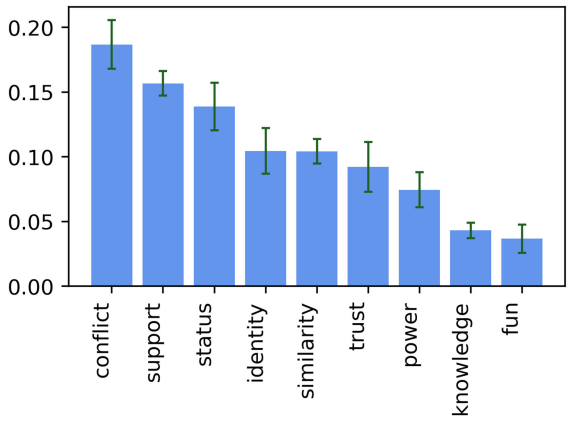

FIGURE 3. Feature importance (absolute value) of the AdaBoost model trained on each type of conversation.

outperform any other individual metric (p-value < 0.05).

- When comparing across three types of analytics (i.e., textual, verbal, and social), we found that the model trained on types of conversations performed the best, achieving an AUC of 76%; followed by the model trained on verbal analytics, and then by the model trained on textual analytics. The model trained on verbal analytics (called *all (V)* in Figure 1 and incorporating prosody, speech rate, pitch, and verbal sentiment) achieved a close second best prediction performance, obtaining an AUC of 74%, which was not statistically significant different from the social analytics model. The model based on all the textual analytics (*all (T)* in Figure 1) performed the worst overall, even if it achieved an AUC as high as 70%.
- By combining all textual, verbal, and social analytics (*All* in Figure 1), the best performing model achieved an AUC of 80% (significantly outperforming any other individual or combined metric), demonstrating that these analytics are, to a certain extent, complementary to each other.

Analysis of Types of Conversations. As the types of conversations were collectively found to be most predictive of meeting success, we then set out to determine which types tended to be more so individually.

First, we inspected the distributions of the types across all meetings (see Figure 2). As one expects, meeting participants mostly exchange *knowledge*, with expressions of shared interests (*similarity*), and a sense of belonging to the same group (*identity*).

Second, we inspected the feature importance of the best performing AdaBoost model trained on types of conversations (see Figure 3). This allowed us to understand which types (positively or negatively) contributed the most to the prediction accuracy. The feature importance is calculated using the standard Gini importance method as the total decrease in AdaBoost tree node impurity (weighted by the probability of reaching the tree node at hand, which is approximated by the proportion of samples reaching that node) averaged over all trees of the ensemble. We found that conflict, support, and status were the most predictive types. *Conflict* usually indicated the presence of diverging views within the meeting, and, eventually, conflict resolution; *support* was associated with emotional or practical aid and companionship; and *status* conferred status, appreciation, gratitude, or admiration upon another. For example, as shown in Table 1, it is not surprising that meetings that provided *support* were considered successful (e.g., *"[...]* very interested to hear about these experiments"). Contrary to conventional wisdom, we observed that *conflict* contributed positively to meeting success. This is partly explained by language exchanges that were mostly constructive, resulting in the definition of common goals or concrete action points (e.g., *"[...] But, let me explain.* [...] You are right, it must be that way".

Overall, these results show that these types of conversations are instrumental to a meeting's success.

DISCUSSION AND CONCLUSION

While existing meeting tools translate, to a great extent, key aspects into analytics, we showed that there exist types of conversations that host important information linked to a meeting's success. If captured, these types of conversations could potentially enrich

TABLE 1. Examples excerpts extracted from the meeting transcripts, illustrating the use of language in the types of conversations under study. Names appearing in the original dialogues were paraphrased, and quotes in boldface indicate language markers concerning corresponding conversation types.

Dimension	Examples
Conflict (contrast, diverging views, conflict resolution)	"Yeah, it's a problem. I think that's like *[...]* I don't know. **But, let me explain**. If you go on Instagram, you push that little heart button, and it starts floating up stuff. That's equivalent when you first get your phone. It's annoying, you know the vibration that most people turn off over time, right? *[...]* they [users] didn't have faith in the buttons being pushed, and over time, you know, maybe then you see the visual feedback. **You are right, it must be that way.**"
Support (giving emotional or practical aid and companionship)	"Human decisions about driver-less or autonomous cars is a very depressing topic. *[...]* The team, though, is **very interested to hear about these experiments**. Welcome everyone." "Thank you for the intro, and thank you for inviting me here. Can I start, right? Yeah, I'm going to talk about the moral machine experiment and a couple of *[...]*" "What about the experiments, and the **follow-up work with these amazing collaborators**?."
Status (appreciation, gratitude, admiration)	"If you have any interesting projects, topics, or ideas that you want to present, please don't hesitate *[...]*, just get in touch." "Exactly. Thank you Daniel, and thank you everyone for participating. See you soon to our next seminar series. Thank you. Thank you for joining us today. Thank you. Thanks for participating"
Identity (shared sense of belonging)	"I mean, but it's certainly one thing that people find interesting to talk about, and **they feel that they have something to share**. I think it's an important thing. Is that supposed to provide people with the opportunity to **give an opinion no matter how crazy or biased they are** *[...]*, but if they could give that opinion."
Similarity (shared interests, motivations or outlooks)	"**From your anecdotal example, it shows one amazing way** that artificial intelligence is used in healthcare. So I think this example is interesting for a couple of reasons. Firstly I think it really well illustrates the potential benefits we could have for medical AI. It also illustrates some of the high-stakes ethical decision-making that these kind of systems would end up being involved in."
Trust (will of relying on the actions of another)	"I don't know, if you guys have any comment on that." "Just one suggestion. Also, I would like to have some project updates. There is a lot of fuzziness around what the project entails, and we have not registered yet. But we don't know what the customer wants to. **We will figure it out, though.**"
Power (having power over the behavior of another)	"These people tend to have the **grace of God**." "*[...]* because of the way that our economy declines, we can now identify specific cultural dimensions."
Knowledge (exchange of ideas learning or teaching)	"**Let me start with an example**. *[...]* you have a typical prediction problem, and that is going to be used in a life-changing decision. *[...]*" "In the final part of the system, we will demonstrate how analog processing works *[...]*. **Any thoughts?**"
Fun (experiening leisure, laughter or joy)	" So, you guys here these sounds? *[...]* Not really in my side. *[...]* Oh, yeah. **It's like bird sounds**. I cannot here hear you on the bridge [referring to WebEx], but I hear voices like birds. **It's so funny**."

meeting analytics, both in real-time and postmeeting. Our results reaffirmed previous findings[18] according to which verbal features (e.g., prosody and pitch) were found to complement textual sentiment and vocabulary ones. Interestingly, we found that certain types of conversations (e.g., conflict and support) were more predictive of meeting success than verbal features, which were the close second best predictive features. In addition, both features were complementary to each other, and a combination of both was more predictive than what they were individually.

Our work offers two main practical implications. First, our types of conversations could be theorized in the context of meetings, and widely adopted in organizational and management research. For example, these types could be linked to the concept of psychological safety. As Edmondson[11] stated, psychological safety refers to "the absence of interpersonal fear that allows people to speak up with work-relevant content." As such a possibility greatly matters in meetings, if captured, it could help teams create safe environments. Second, our models could be deployed

and integrated with any communication tool that provides voice recordings. MeetCues[2] is an example of such a tool: it allows participants to engage during a meeting, and to reflect on their experience through visual and interactive features.

This work has three main limitations that call for future research. First, our dataset refers to business meetings, thus our findings might not generalize to other types of meetings. In addition, given our relatively small dataset, we used a leave-one-out validation procedure and, as such, our models might have learned intrinsic patterns from the participants rather than from the actual meetings, questioning the generalizability of these models. Future work should consider to: (i) apply the models to other types of meetings and companies, testing their generalizability; and (ii) build prediction models that are more fine-grained than ours, which were based on dichotomized success scores. Second, we adopted audio as our main source. However, other aspects derived from facial expressions or body languages might be able to capture more nuanced emotions from meeting participants (e.g., key turning points in a meeting).[4] Finally, our types of conversations capture the most frequent dynamics of interpersonal exchanges in general settings, which are not specific to meetings. Further tailoring those conversation types to the meeting context might boost the model performance, pushing it even further beyond our model's fairly high AUC of 76%.

Our work shed light on the importance of quantifying different types of conversations at scale. By monitoring these conversations within an organization (e.g., company, university), one could potentially measure specific aspects of organizational productivity, and proactively take actions for improvement. For example, our analytics can be integrated as a plug-in for monitoring and improving online conference/meeting applications (e.g., Zoom).[2] While this approach promises to improve organizational productivity, it also raises questions related to workplace surveillance.[3] On a very pragmatic level, there is a handful of reasons as to why organizations opt in for surveillance (e.g., maintaining productivity, monitoring resources used, protecting the organization from legal liabilities). Critics, however, might rightly argue that there is a fine line between what organizations could be monitoring and what they should be monitoring. If crossed, this line will have unintended consequences directly on employees, affecting their well-being, work culture, and productivity. If future meetings tools incorporate any kind of monitoring, they would need to ensure that such a monitoring is done in a way that preserves an individual's rights, including that of privacy.

ACKNOWLEDGMENTS

We thank those who actively supported this research at Nokia Bell Labs; in particular, Mark Clougherty, Sean Kennedy, and Marcus Weldon for their guidance during the development of the meeting companion app *MeetCues*; and Iraj Saniee for his support in researching the audio sentiment classifiers.

REFERENCES

1. A. Akbik, T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf, "Flair: An easy-to-use framework for state-of-the-art NLP," in *Proc. 2019 Conf. North Amer. Chapter Assoc. Comput. Linguistics (Demonstrations)*, 2019, pp. 54–59, doi: 10.18653/v1/N19-4010.
2. B. A. Aseniero, M. Constantinides, S. Joglekar, K. Zhou, and D. Quercia, "MeetCues: Supporting online meetings experience," in *Proc. IEEE Visual. Conf.*, 2020, pp. 236–240, doi: 10.1109/VIS47514.2020.00054.
3. K. Ball, "Workplace surveillance: An overview," *Labor Hist.*, vol. 51, no. 1, pp. 87–106, 2010, doi: 10.1080/00236561003654776.
4. J.-H. Choi, M. Constantinides, S. Joglekar, and D. Quercia, "KAIROS: Talking heads and moving bodies for successful meetings," in *Proc. Int. Workshop Mobile Comput. Syst. Appl.*, 2021, pp. 30–36, doi: 10.1145/3446382.3448361.
5. M. Choi, L. M. Aiello, K. Z. Varga, and D. Quercia, "Ten social dimensions of conversations and relationships," in *Proc. The Web Conf.*, 2020, pp. 1514–1525, doi: 10.1145/3366423.3380224.
6. M. Constantinides, S. Šćepanović, D. Quercia, H. Li, U. Sassi, and M. Eggleston, "COMFEEL: Productivity is a matter of the senses too," *Pro. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 4, pp. 1–21, 2020, doi: 10.1145/3432234.
7. E. Coutinho and N. Dibben, "Psychoacoustic cues to emotion in speech prosody and music," *Cogn. Emotion*, vol. 27, no. 4, pp. 658–684, 2013, doi: 10.1080/02699931.2012.732559.
8. M. G. de Pinto, M. Polignano, P. Lops, and G. Semeraro, "Emotions understanding model from spoken language using deep neural networks and mel-frequency cepstral coefficients," in *Proc. IEEE Conf. Evolving Adaptive Intell. Syst.*, 2020, pp. 1–5, doi: 10.1109/EAIS48028.2020.9122698.
9. S. Deri, J. Rappaz, L. M. Aiello, and D. Quercia, "Coloring in the links: Capturing social ties as they are perceived," *Proc. ACM Hum.- Comput. Interact.*, vol. 2, pp. 1–18, 2018, doi: 10.1145/3274312.

10. W. Dong, B. Lepri, T. Kim, F. Pianesi, and A. S. Pentland, "Modeling conversational dynamics and performance in a social dilemma task," in *Proc. 5th Int. Symp. Commun., Control Signal Process.*, 2012, pp. 1–4, doi: 10.1109/ISCCSP.2012.6217775.
11. A. Edmondson, "Psychological safety and learning behavior in work teams," *Administ. Sci. Quart.*, vol. 44, no. 2, pp. 350–383, 1999, doi: 10.2307/2666999.
12. C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 8, 2014, pp. 216–225.
13. V. Këpuska and G. Bohouta, "Comparing speech recognition systems (Microsoft API, Google API, and CMU Sphinx)," *Int. J. Eng. Res. Appl*, vol. 7, no. 3, pp. 20–24, 2017, doi: 10.9790/9622-0703022024.
14. J. Lee and S. C. Marsella, "Predicting speaker head nods and the effects of affective information," *IEEE Trans. Multimedia*, vol. 12, no. 6, pp. 552–562, Oct. 2010, doi: 10.1109/TMM.2010.2051874.
15. W. Morgan, P.-C. Chang, S. Gupta, and J. Brenier, "Automatically detecting action items in audio meeting recordings," in *Proc. 7th SIGDIAL Workshop Discourse Dialogue*, 2006, pp. 96–103, doi: 10.5555/1654595.1654614.
16. G. Murray, "Uncovering hidden sentiment in meetings," in *Proc. Can. Conf. Artif. Intell.*, 2016, pp. 64–72, doi: 10.1007/978-3-319-34111-8_9.
17. I. R. Murray and J. L. Arnott, "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," *J. Acoust. Soc. Amer.*, vol. 93, no. 2, pp. 1097–1108, 1993, doi: 10.1121/1.405558.
18. T. L. Nwe, S. W. Foo, and L. C. De Silva, "Speech emotion recognition using hidden Markov models," *Speech Commun.*, vol. 41, no. 4, pp. 603–623, 2003, doi: 10.1016/S0167-6393(03)00099-2.
19. S. Sarda *et al.*, "Real-time feedback system for monitoring and facilitating discussions," in *Proc. Natural Interact. Robots, Knowbots Smartphones*, 2014, pp. 375–387, doi: 10.1007/978-1-4614-8280-2_34.
20. S. Somasundaran, J. Ruppenhofer, and J. Wiebe, "Detecting arguing and sentiment in meetings," in *Proc. 8th SIGDIAL Workshop Discourse Dialogue*, 2007, pp. 26–34.

KE ZHOU is a Senior Research Scientist in the Social Dynamics team at Nokia Bell Labs Cambridge, U.K. and an Assistant Professor of computer science at the University of Nottingham. His research interests include information retrieval and user modeling. He received the Ph.D. degree from the University of Glasgow, Glasgow, U.K. He is the corresponding author of this article. Contact him at ke.zhou@nokia-bell-labs.com.

MARIOS CONSTANTINIDES is a Research Scientist in the Social Dynamics team at Nokia Bell Labs Cambridge, U.K. His research interests include human–computer interaction, mobile-sensing, and user modeling. He received the Ph.D. degree from University College London, London, U.K. Contact him at marios.constantinides@nokia-bell-labs.com.

LUCA MARIA AIELLO is an Associate Professor at the IT University of Copenhagen (DK). He conducts interdisciplinary computational social science research. He received the Ph.D. degree from the University of Turin, Turin, Italy. Contact him at luai@itu.dk.

SAGAR JOGLEKAR is a Research Scientist in the Social Dynamics team at Nokia Bell Labs Cambridge, U.K. His research interests include representation learning and its practical applications in the fields of social computing and urban informatics. He received the Ph.D. degree from King's College London, London, U.K. Contact him at sagarjoglekar@gmail.com.

DANIELE QUERCIA is the Department Head at Nokia Bell Labs in Cambridge, U.K. and a Professor of urban informatics at King's College London, U.K. His research interests include computational social science and urban informatics. He received the Ph.D. degree from University College London, London, U.K. Contact him at quercia@cantab.net.