Theme Article:
Conversational User Interfaces and Interactions

Towards Interpersonal Assistants: Next-Generation Conversational Agents

Opportunities With Always-on Microstructural Conversation Intervention Assistants

Inseok Hwang
IBM

Youngki Lee
Seoul National University

Chungkuk Yoo
IBM

Chulhong Min
Nokia Bell Labs

Dongsun Yim
Ewha Womans University

John Kim
KAIST

***Abstract*—We propose novel interpersonal assistants, a next-generation conversational agent which is always-on, unobtrusively serving natural human-to-human conversations. We deepen the motivation and design insights with real practices in language delays and parent–child conflicts. We then present a common platform initiative to effectively support rapid development of interpersonal assistant applications, with a highlight on the key functional element of turn isolations and technical insights on microstructural dynamics.**

Conversation would be the most prevailing mode of social interaction. Recently, we have witnessed rapid penetration of intelligent services that listen and respond to human conversations in everyday living space and even on-the-go. Commercial services such as Amazon Alexa, Apple Siri, and Google Assistant are becoming increasingly pervasive. Most of today's services

Digital Object Identifier 10.1109/MPRV.2019.2922907
Date of current version 25 July 2019.

1536-1268 © 2019 IEEE

share a common design metaphor—a personal assistant. An intelligent service naturally serves the user by acting as a conversation counterpart.

Imagine what would be the next generation of intelligent services beyond a personal assistant. Among many possibilities, we propose an *interpersonal assistant* for human-to-human conversation. Rather than inheriting a conventional model between a person and a computer, an interpersonal assistant would serve a pair or a greater number of people interacting together, monitoring the on-going conversation unobtrusively and offering useful services just-in-time. For example, the assistant facilitates conversation to keep going well,[1] helps note-taking of an on-going conversation,[2] or provides an impromptu intervention when it spots potentially risky interaction.[3,4]

We note that commercial voice-enabled assistants are being widely appropriated for interaction coaching and training,[5,6] mainly through an active conversation between a human user and a machine assistant. We highlight that our interpersonal assistants discussed in this paper are uniquely positioned differently from such assistants. Our interpersonal assistant sits between a group of face-to-face conversing human users. The interpersonal assistant is not an active interaction counterpart most of the time; it remains unobtrusive while carefully monitoring the human-to-human interaction, and makes a gentle, application-specific suggestions when appropriate. Also, our interpersonal assistant allows same-time application of such suggestions immediately in the on-going conversation, unlike the existing voice-enabled assistants letting the user learn first in an emulated setting and later apply the lesson in a real situation.[5,6]

The key contribution of this paper is to present a newly emerging form of conversational agents, namely interpersonal assistant. In particular, our discussion encompasses the vertical integration of our two separately developed layers proposed previously—the exemplary interpersonal assistant application grounded in clinical domains[3] and the low-power, low-latency continuous conversation monitoring technology for mobile devices,[7] on which brief summaries are provided in this paper. We also present a novel horizontal expansion to a newly explored application that potentially benefits those in early parenthood experiencing hard conflicts with their young children. Notably, we present a two-dimensional (2-D) model through which we characterize the distinct patterns of parent–child conflicts and derive the computational inferability. Furthermore, our discussion expands to a more usable and versatile interpersonal monitoring platform incorporating mobile, wearable, and home IoT environments. Through the top–down discussion, we identify the prominent properties of *nonverbal microstructural dynamics* that interpersonal assistant applications share in common, allowing the applications to benefit from rapid, systematic development with uniform platform supports.

In the remainder of the paper, we first deep-dive into two compelling applications enhancing parent–child conversations. Leveraging the microstructural dynamics properties that we identified from the applications, we discuss the platform initiative to facilitate developing interpersonal assistants with high computational efficiency. Then, we present a set of pilot experiments that exhibit the feasibility and usability of our exemplary interpersonal assistants that we have discussed.

DEEP-DIVES: PARENT–CHILD CONVERSATIONS

Parents are encouraged to have lots of interaction with their young children. It is agreed that making attentive, thoughtful interactions with a child is an integral element in parenting. We outline two sets of our in-depth studies on certain types of parent–child conversations from clinical and developmental perspectives.

Study I: Language Delay

In speech–language pathology, language delay broadly describes developmental problems in young children not acquiring language as expected for their chronological age.[8] Lifetime impacts of persistent language difficulties can be severe as it can lead to poor social relations, literacy impairment, and even low socioeconomic status.[9]

Speech–language pathologists (SLPs) established therapeutic procedures for children with language delay. Those include in-clinic procedures with structured protocols and out-of-clinic guidelines for prolonged continuing care in daily life. In-clinic procedures are typically a structured interaction session between an SLP and a child. A

Table 1. Parent training guidelines frequently given by SLPs.

Classification	Guidelines to the parents
Microstructural	"*Talk more slowly.*"
	"*Wait for the child to talk back.*"
	"*Do not interrupt the child before she completes what she says.*"
	"*Talk in short (and simple) sentences.*"
	"*Respond immediately when the child talks first.*"
	"*Make more turn-takings with the child.*"
	"*Spend more time talking with the child.*"
	"*Articulate what you speak.*"
Semantic	"*Praise the child.*"
	"*Set a topic that the child is interested in.*"
	"*Use positive words.*"
	"*Refrain from making one-sided instructions.*"
	"*Repeat the important keyword.*"
Facial	"*Make eye-contact with the child.*"

representative example of out-of-clinic guidelines is *parent training*,[10] which trains the parents, who are the dominant interaction counterpart of the child, to be active and hands-on facilitators of their child's language development.

In prior work,[3] we identified the key findings on how the parent training is exercised in the wild, through extensive user studies with 8 SLPs and 13 parents treating their children with language delay. A new parent training typically begins with an SLP observing a parent and a child interacting together. Then, the SLP provides selected guidelines curated for their own interaction styles. Table 1 lists the representative guidelines that the SLPs frequently provide. Although those guidelines look straightforward and the parents know every guideline well, it turned out that the parents were experiencing severe difficulties to abide by the guidelines in their daily conversations. Notably, we found that the fundamental challenge lies in the parents' cumulative mental fatigue in taking care of a child with language delay for a prolonged period, yet with almost no perceivable progress week after weeks. To many parents, it was hardly possible to refrain from momentary temper tantrums, recall the proper parent guideline, and apply that guideline in a second. Moreover, many of parent guidelines are against their life-long developed speaking habits. To simply put, we witnessed their real-life hardship rooted in the very human nature susceptible to overwhelming emotion and chronic fatigue, as well as old habits dying hard.

Study II: Parent–Child Conflicts

Raising a child is likely a continuation of day-to-day conflicts and negotiations. As the child begins interacting with their parents, conflicts are inevitable for mundane reasons such as eating, cleaning, dressing, and sleeping. We interviewed ten parents with young children. We found that they got exhausted from the same tug-of-war every day. One mother stated, "*It is so exhausting to wrestle with my child every morning. I always have to get into a fight about brushing teeth and having breakfast.*" Parenting strategies in the moment of parent–child conflicts largely affect the development of the child's ability to regulate emotions. Children learn from getting their demands rejected in conflicts with parents. They also learn how to negotiate therefrom. This ability is one of the essential factors in building their life-long styles of interacting with others.

Parent effectiveness training (PET) highlights the importance of parents' interaction styles, strategies, and attitudes as cornerstones toward smooth resolution of parent–child conflicts.[11] For an in-depth understanding, we interviewed a pediatric psychiatrist and four developmental psychotherapists. They considered nonverbal responses such as facial expressions and voice tones as a key to resolve a conflict in a constructive way. According to them, children whose language development is in progress are highly sensitive to nonverbal messages from their parents. Parents should avoid conveying negative emotions and aggressive messages such as frowning, giving threatening looks, or yelling at their child. All the experts reported that parents do not often recognize their exposure of negative emotion and nonverbal messages. The pediatric

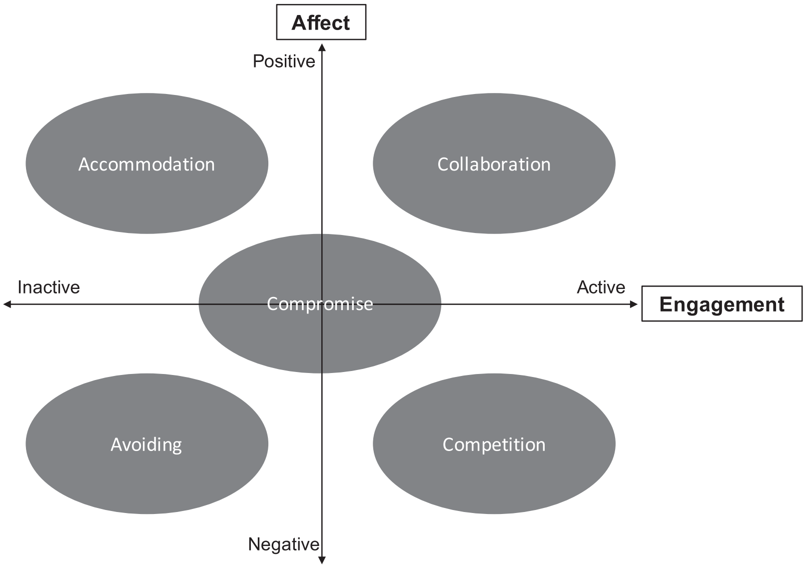

Figure 1. Two-dimensional model of conflicting interactions.

psychiatrist mentioned, "*They manage to control their verbal messages, but they fail to self-regulate their nonverbal messages.*" In our parental interviews, one mother said, *"We went to a restaurant. (…) Finally I ended up shouting at him. (…) My husband said he saw madness in my eyes. I turned into someone like a lunatic, and flipped out. I didn't realize that I lashed out my child so badly."* It is not straightforward for parents to be aware of and change their interaction styles that have been established and habituated for a long time.

To facilitate modeling of the negativity in nonverbal messages in parent–child conflicts, we first refer to an organizational study that elicited five major strategies appearing in conflicts: *avoiding*, *competition*, *compromise*, *accommodation*, and *collaboration*.[12] We also refer to the conflict coding scheme,[13] which uses the two orthogonal dimensions in coding, *affect,* and *engagement*. The affect dimension represents the extent of friendliness in one's emotional response. The engagement dimension represents the degree of one's participation in a conflicting interaction. Then, we attempt to position those strategies on the 2-D plane as shown in Figure 1.[12–14] For example, the collaboration strategy seeks for a resolution to satisfy both parties, which naturally involves active engagement and positive affect. The competition strategy pursues exclusively one's own satisfaction while having the other give up, which would most likely involve active engagement and negative affect. The accommodation and avoiding strategies do not likely involve active engagement; the difference is that the accommodation seeks for a resolution by one voluntarily accepting the other's demands, while the avoiding attempts to stop the conflict by ignoring the other. In the parent–child conflict context, we postulate that the negative and aggressive messages would be closely related to the avoiding or competition strategies in the 2-D model, while specific guidelines in response to these two strategies may differ in between. In the later section, we demonstrate that the degrees of affect and engagement are computationally inferable by monitoring nonverbal features such as turn-takings and prosody in interpersonal conversation.

24/7 Interpersonal Assistant to Expedite Competence

Imagine an interpersonal assistant running on mobile, wearable, or home IoT devices always available to the parent–child pair. It will cover their daily routines, continuously listen to most of the conversations between them, spot a moment when the parent is likely straying from one of the guidelines, and provide a timely nudge to the parent, so that the parent recalls the guideline and correctly applies it. It would offer user experiences analogous to being with a professional who observes their daily routines and advises them just-in-time.

Recent advances in natural language processing shed light on the feasibility of such an interpersonal assistant for parent training. Spurred by novel architectures based on deep neural networks, semantic understanding on natural language sentences has attained promising performances.[15] Commercial services such as IBM Watson and Microsoft Cognitive Services offer building blocks for conversational services to listen, classify, and respond to a speech.[16] However, building a general interpersonal assistant supporting children's conversation is not straightforward. Recognition on young children's speech exhibits far less performance than adults' speech,[17] mainly due to children's underdeveloped pronunciation, ungrammatical phrases, frequently changing speech characteristics as they grow, and insufficient speech corpus specialized in children.

Our key insight for a breakthrough was the dominance of microstructural guidelines outnumbering semantic or facial guidelines. We define "microstructural" as the fine-grained temporal dynamics of interpersonal or intrapersonal conversation such as the speakers' turn-takings, turn-durations, turn-overlaps, the phonemic rate in a speaking turn, etc. An average turn duration is only about 2 s and a minimum turn duration is as short as 300 ms.[7] Importantly, detecting a moment when the conversation strays from one of the microstructural guidelines does not necessarily require verbal properties nor semantic understanding, eliminating the speech recognition poorly performing with young children.

For example, SLPs usually give the guideline of "*Wait for the child to talk back*" when the parent repeats questions before the child responds. Our assistant would determine such moment by spotting an on-going turn-taking pattern that the parent's turns appear several times without the child's turn in the middle. Similarly, for the guideline of "*Do not interrupt the child before she completes what she says*," the parent may be nudged when the parent's turn cuts off the child's on-going turn. Once our assistant segments the conversation into individual turns, the assistant would further analyze each turn to extract nonverbal phonemic features to determine per-turn speech rates, so that it triggers the guideline of "*Talk more slowly*" accordingly.

We note that this insight on the significance of microstructural and nonverbal properties seamlessly applies to the parent–child conflict scenarios as well. First, the parenting experts highlight the regulation of negative emotion and nonverbal messages in parent–child conflicts given the premature verbal development of young children. Also our particular interest in the emotional element makes the nonverbal channel more significant, which is known to account for 93% in conveying emotions.[18] Second, the use of fine-grained nonverbal properties would be promising to determine the on-going strategy in conflicts, as reported that each of the five conflict strategies exhibits its own unique trends in nonverbal properties.[19] The competition strategy exhibits higher and sharper vocal tones, speedy utterances, excessive physical gestures, and head-shakings; the avoiding strategy exhibits longer silence and avoidance of eye-contacts. In contrast, the collaboration strategy exhibits longer turn-durations, slower utterances, smooth, and stable vocal tones. When our interpersonal assistant detects the elevation of a negative conflicting strategy, it intervenes in the conversation so that a parent can regulate his/her emotion and try to constructively resolve the conflict regarding the assistant's guidelines.

For example, we note the psychotherapists advise the parent not to use negative nonverbal messages, e.g., blaming, yelling, or being sarcastic. Our interpersonal assistant detects conflictual turns from alternating high/sharp voice of the parent and crying child. The assistant then delivers the guidelines of "*Don't be angry at your child*" and/or "*Listen what your child says.*" Psychotherapists also advise to realize how the parents themselves look upon a conflict. Our interpersonal assistant would extend to have a wearable camera attached on the child's clothes and give a guideline of "*Keep a friendly face.*" It can take a video of the parent upon a conflictual moment so that the parent reviews their reaction in real time through a wearable eyewear or later with a psychotherapist.

Despite many parameters to be empirically determined, we believe that the prominence of microstructural properties in the context of healthier parent–child interaction would make a great leap toward realizing a just-in-time interpersonal assistant in the very near future.

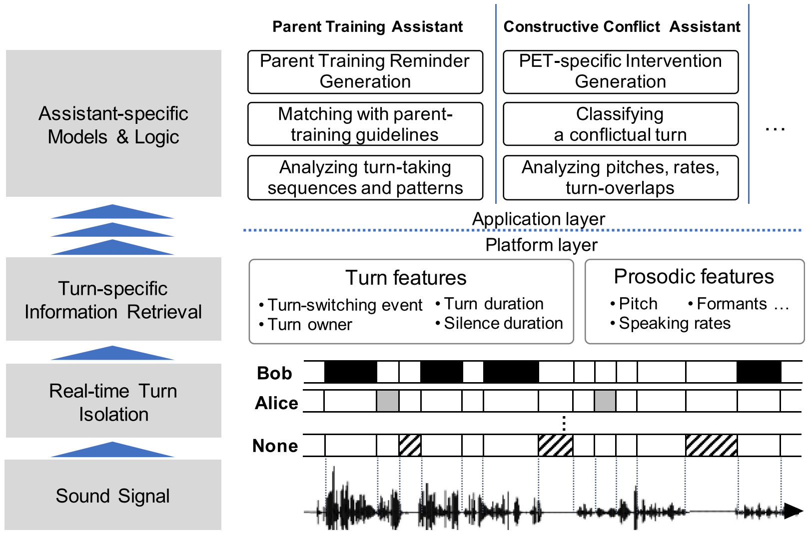

Figure 2. Vertical layered architecture for interpersonal assistants, consisting of the common supporting modules belonging to the underlying platform layer and the individual assistant-specific modules on top of them.

PLATFORM INITIATIVE

We discussed exemplary real-world motivations for interpersonal assistant services and their key properties in favor of technical feasibility. We note the prominence of microstructural properties is not limited to a particular example of interpersonal assistants but also applying to virtual agents helping people improve social communication skills.[6] Those virtual agents could be reinvented as interpersonal assistants so that the user can benefit from just-in-time feedback.

An ideal interpersonal assistant would feature a comprehensive coverage throughout one's daily life. These objectives lead to high-level technical requirements:

- Real-time processing of on-going conversation to ensure just-in-time intervention.
- Runnable on commodity mobile or wearable devices that stay with the user as long as possible.
- Highly energy-efficient to ensure always-on operation throughout a daily routine.

The future potential of interpersonal assistants and the nontrivial technical requirements strongly motivated us to seek a vertical layer separation approach, i.e., a well-defined underlying platform supporting various interpersonal assistants efficiently. The platform would offer a suite of APIs to lessen the efforts in developing interpersonal assistants and provide a framework that takes over the runtime complexities in processing speech signals and managing computing resources, as depicted in Figure 2.

The key primitive APIs would be those listening to an on-going conversation, isolating individual turns from the conversation, and retrieving per-turn information, e.g., the speaker, the duration, the speaking rate and prosody, the transcript, etc. An interpersonal assistant application running on top of this platform can easily subscribe to a continuous stream of per-turn information from an on-going conversation. The interpersonal assistant would have the information go through its application-specific logic, spotting a right moment to issue a service action.

Microspatial Multidevice Auditory Fusion Network

What if we think of the emerging environments of multiple devices, multiple wearables per user? What if such wearables have evolved into an open system capable of accommodating multiple third-party applications, just like our phones and watches? There are already numerous tiny wearable devices available in the market or crowd-funding projects. We can reasonably assume multiple devices concurrently available with a pair of conversing users in near future.

Leveraging such environments, we envision Microspatial Multi-device Auditory Fusion

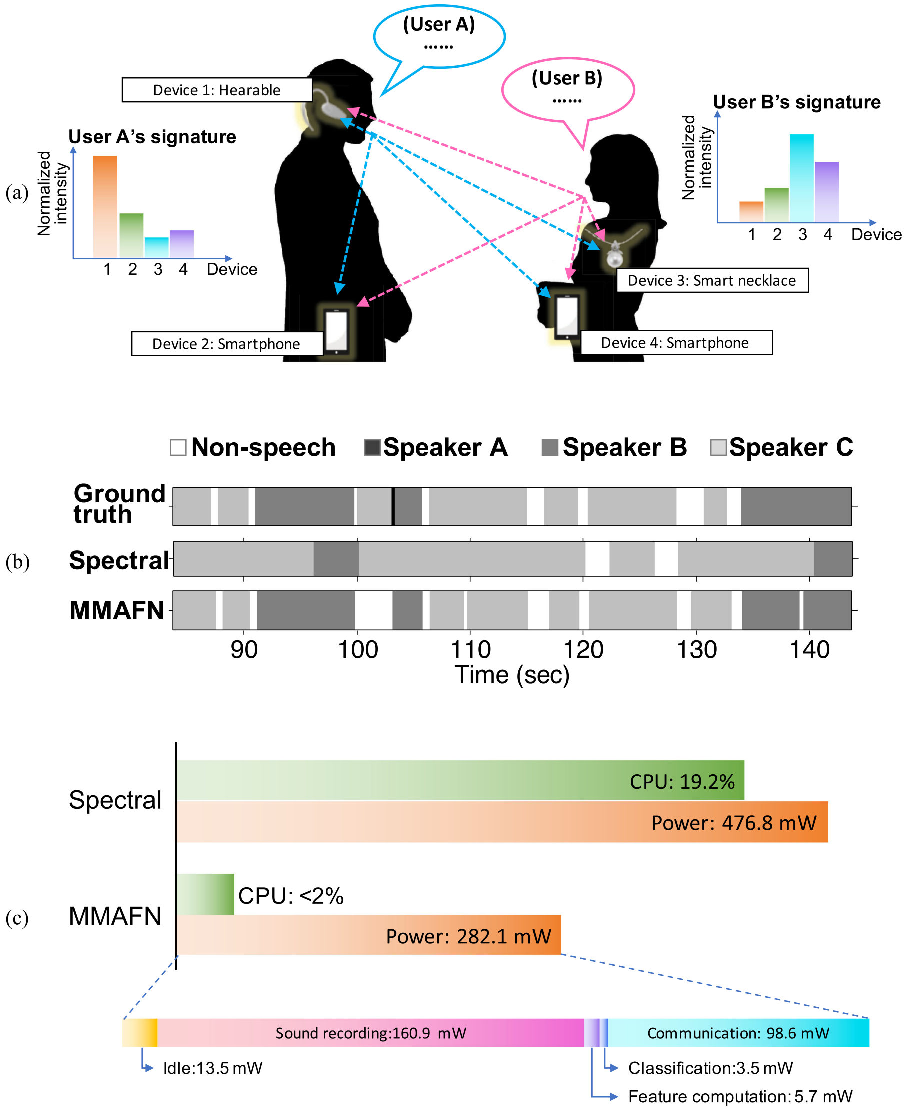

Figure 3. Illustration and performance of MMAFN techniques. (a) MMAFN technique principles. (b) Turn-monitoring accuracy on real-world conversations. (c) CPU utilization and power consumption.

Network (MMAFN) to realize accurate, sustainable platform support for 24/7 interpersonal assistants. Given a conversing pair, we note the spatially distributed property of the devices each of which is located likely at a unique body location of either person. Then, we have all such devices collaboratively form a 3-D distributed network covering a bubbly space surrounding the conversing pair. Due to the spatially distributed nature, each device would have distinct sound transmission paths from either person's vocal cord to the device, where the path may be through air, bone, or both. When either person is speaking, the voice propagating from the person's vocal cord to each device should undergo a certain rate of attenuation governed by the distance and medium through the device-specific transmission path. The key insight of MMFAN is that, as long as a device's location changes minimally, the device-specific attenuation rate would be invariant to the original voice intensity produced at a person's vocal cords. While the device itself is unable to determine the original intensity, comparing the received intensities at other devices should reveal constant relative device-specific attenuation rates. In other words, a normalized vector of the voice intensities received at a set of devices will be a person-specific signature in a stable conversation setting as illustrated in Figure 3(a). This signature remains invariant to the momentary variation of the person's speaking loudness, speed, or phonemes. The principles of MMAFN apply to not only a pair but also to an arbitrary number of conversing people.

In our prior work, we built an initial version of MMAFN working on a group of smartphones distributed around a conversing group.[7] Figure 3(b) demonstrates a part of a turn-taking diagram observed from a triadic conversation with three devices available. While the ground-truth turn-taking diagram includes many short turns and pauses, the control group using the existing spectral feature-based speaker identification (denoted Spectral[20]) fails to capture most of the short turns and pauses, resulting in incorrectly merged turns. MMAFN successfully captures many of them. The overall turn-monitoring accuracy from three sessions of 15-min unconstrained conversations was 80.5% for Spectral and 93.0% for MMAFN, reducing the failures by 64.1%. Figure 3(c) shows that MMAFN outperforms Spectral in terms of the resource efficiency.

Our MMAFN stack identifies real-time turn-taking patterns and timing parameters such as time gap between turns, duration of overlaps, dominance, and so on. On top of this stack, we developed an additional layer capable of computing more than 100 nonverbal vocal features per turn. These features enable our assistant to perform in-depth nonverbal analysis on individual turns.

Breaking down the power consumption reveals a further power-saving potential; the sound recording dominates the overall power consumption, and the core computations in MMAFN hardly draws power. Extensive research efforts are being put to offload repetitive sensor data processing to a small, power-efficient processor. Constant sound-recording would be a great candidate for such optimization.

PILOT EXPERIMENTS

We developed two pilot interpersonal assistants. For providing just-in-time intervention to a parent talking to a child with language delay, an intervention is given using one of the five verbal reminders listed in Table 1 when the parent is straying from a guideline. For nudging a parent in a conflict with a child, a notification is sent to the parent when a conflict situation has been elevated.

Interpersonal Assistant: Child With Language Delay

We deployed the interpersonal assistant on three clinical protocols. Each session was a 20-min free-form parent–child conversation remotely observed by an SLP. The conversations were recorded and transcribed, from which the SLP determined a moment, on her own discretion, that the parent needed to be reminded of a certain guideline. We had the interpersonal assistant running in real time but muted the verbal reminder from the assistant, in order not to interfere with the clinical protocols. We cross-validated the two sets of reminder logs—those generated by the interpersonal assistant and those determined by the SLP. For all three sessions, the SLP determined 33 reminders necessary while the assistant generated 31 reminders. Among them we identified 22 matches, 9 false-positives, and 11 false-negatives, which yield the overall 71.0% precision and 66.7% recall. Major reasons of false-positives included the parents' facial responses that the SLP deemed a legitimate replacement of a verbal response. False-negatives are mainly caused by the parents' varying speech rates within a turn. Although the average speech rate during the turn was not high, the SLP spotted a peak. We discussed more detailed analysis in the previous work.[3]

Interpersonal Assistant: Parent–Child Conflict

We conducted a set of data-driven analyses to evaluate the performance of MMAFN in the context of parent–child conflicts. Unlike the earlier experiment in the context of language delay, this experiment was performed with a corpus of precollected dataset.[14]

We developed a statistic model that infers the emergence of negative affect and/or inactive engagement. We used a corpus of 50-h-long videos prerecorded for child development counseling purposes, consisting of daily life scenes of conflicting interactions between family members including young children. Manually segmenting individual turns from these videos resulted in

Table 2. Correlation coefficients between nonverbal features and conflict strategy dimensions.

Features from MMAFN		Pearson's correlation coefficient (*: p < 0.003, ** < 0.001)	
		Affect	Engagement
Turn features	Dominance (ratio of one's speaking turn duration)	-0.029	0.321**
	Response time (avg. speaking turn interval)	0.080	-0.151 *
	Overlap (# of overlapped turns' duration)	-0.415**	0.263**
	Awkward silence (# of silence turns' duration > 3 secs)	0.054	-0.248**
	Sparseness (speaking turns' duration per minute)	-0.101	0.356**
	Interactivity (# of speaking turns per minute)	0.267**	-0.129
Prosodic features	F0 (fundamental frequency)	-0.579**	0.487**
	Jitter (cycle-to-cycle variation of fundamental frequency)	0.122	-0.222**
	Shimmer (amplitude deviations between pitch period)	0.138 *	0.206**
	Energy (root-mean-square signal frame energy)	-0.242*	0.289*
	Loudness (normalized frame intensity)	0.180**	0.224**

971 turns. For each turn, 11 trained human annotators scored the level of affect and engagement in a 7-point scale, respectively. We then considered the average score for each turn as the ground-truth. We computed the correlation coefficients between the human-coded affect/engagement scores and various nonverbal features that MMAFN supports. Among more than 100 features, we list a few major features in Table 2. The F0 feature exhibits the strongest correlations in both affect and engagement, followed by the overlap feature in affect and by the sparseness feature in engagement, respectively. These findings are consistent with the literature that reported the signal characteristics observed from each conflict strategy.[19] We built a regression model inferring the affect and engagement scores, respectively. With six regression methods (Random Forest, Mean, KNN, Linear, PLS, Support Vector Regression) and tenfold cross validation, the Random Forest performed best for inferring the engagement score (0.67 RMSE) and the Linear performed best for the affect score (0.43 RMSE).

Besides the evaluations above, we conducted interviews with 20 parents about our prototypes, i.e., 10 parents for the language delay assistant and 10 parents for the parent–child conflict assistant. Most of the parents agreed that the assistants would help them get used to the guidelines about their interaction styles. Many parents pointed out that prolonged use of the assistant would be an important factor to acquire desirable changes of their conflict-prone tendency. As for the parent–child conflict assistant, one mother stated: "*Children keep growing and their demands also keep changing. I had different conflicts due to different reasons. I think I need to keep viewing myself constantly.*" Some parents had a concern that pointing out what they are wrong might hurt, so that they might be reluctant to receive the intervention. For the concern, one parent suggested, "*It's like the carrot and stick. Stick this time. Then, a mom needs a carrot.*" In this light, our assistant can monitor progress of a detected conflictual moment and deliver a compliment if the parent resolves the conflict in a peaceful way. Another discussion with parents was a privacy concern about the audio recording. Most parents did not care about it because they had serious motivations to use the assistants. We note that monitoring microstructural properties with MMAFN does not require high quality audio recording, e.g., only 500 Hz of sampling rate in audio recording was enough to extract microstructural properties[7] at which frequency recovering the linguistic contents is impossible.

CONCLUSION

We outlined interpersonal assistants as a promising model that conversational agents may evolve. Through two sets of informed studies with clinical, psychological, and developmental experts, we demonstrated the compelling potential of such interpersonal assistants. Furthermore, we elicited key functional elements for always-on services running on resource-scarce devices. We anticipate that our deep-dives and platform initiative are meaningful steps toward a holistic framework spurring the emergence of new instances of interpersonal assistants.

ACKNOWLEDGMENT

The authors would like to thank our colleague Changhun Lee for his invaluable help to collect the parent–child conflict dataset and conduct preliminary analysis on the data.

REFERENCES

1. M.-H. Lee *et al.*, "Flower-Pop: Facilitating casual group conversations with multiple mobile devices," in *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, 2017, vol. 1, no. 4, Art. no. 150.
2. V. Kalnikaitė *et al.*, "Markup as you talk: Establishing effective memory cues while still contributing to a meeting," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, 2012, pp. 349–358.
3. I. Hwang *et al.*, "TalkBetter: Family-driven mobile intervention care for children with language delay," in *Proc. ACM Conf. Comput. Supported Cooperative Work*, 2014, pp. 1283–1296.
4. S. Song *et al.*, "TalkLIME: Mobile system intervention to improve parent-child interaction for children with language delay," in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 304–315.
5. A. Danielescu and G. Christian, "A bot is not a polyglot: Designing personalities for multi-lingual conversational agents," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2018, Paper no. CS01.
6. A. Pradhan *et al.*, "Accessibility came by accident: Use of voice-controlled intelligent personal assistants by people with disabilities," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2018, Paper No. 459.
7. Y. Lee *et al.*, "SocioPhone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion," in *Proc. ACM MobiSys*, 2013, pp. 375–388.
8. R. Paul *et al.*, *Language Disorders from Infancy through Adolescence: Listening, Speaking, Reading, Writing, and Communicating*, Amsterdam, The Netherlands: Elsevier, vol. 4e, 2011.
9. J. Clegg *et al.*, "Developmental language disorders: A follow-up in later adult life. cognitive, language, and psychosocial outcomes," *Child Psychol. Psychiatry*, vol. 46, no. 2, pp. 128–149, 2005.
10. H. A. Manolson, "Parent training: A means of implementing pragmatics in early language remediation," *Human Commun. (Summer 1979)*, vol. 4, pp. 275–281, 1979.
11. T. Gordon, Parent Effectiveness Training: The Proven Program for Raising Responsible Children. Potter/Ten Speed/Harmony/Rodale, New York, NY, 2008.
12. K. W. Thomas *et al.*, "Conflict and conflict management: Reflections and update," *J. Organizational Behav.*, vol. 13, no. 3, pp. 265–274, 1992.
13. A. L. Sillars, "Procedures for coding interpersonal conflict: The verbal tactics coding scheme (VTCS)," Dept. Commun. Studies, Univ. Montana, Missoula, MT, USA, 1986.
14. C. Lee, "Computational model of conflictual interactions between people in close relationship," *Master's Thesis, Div. Web Sci. Technol.*, KAIST, Daejeon, South Korea, 2015 (In Korean language).
15. Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2014, pp. 1746–1751.
16. J. Lee *et al.*, "Sci-Fii: Speculative conversational interface framework for incremental inference on modularized services," in *Proc. 18th IEEE Int. Conf. Mobile Data Manage.*, 2017, pp. 278–285.
17. J. Kennedy *et al.*, "Child speech recognition in human-robot interaction: Evaluations and recommendations," in *Proc. ACM/IEEE Int. Conf. Human–Robot Interact.*, 2017, pp. 82–90.
18. A. Mehrabian, Silent Messages. Belmont, CA, USA: Wadsworth, 1971.
19. L. K. Guerrero *et al.*, "Conflict and disengagement," Nonverbal Communication in Close Relationships. New York, NY, USA: Taylor and Francis, pp. 198–224, 2005.
20. H. Lu *et al.*, "SpeakerSense: Energy efficient unobtrusive speaker identification on mobile phones," in *Proc. Int. Conf. Pervasive Comput*, 2011, pp. 188–205.

Inseok Hwang is a research staff member at IBM and an IBM Master Inventor. His research interests include mobile/IoT systems, embedded AI, and human–computer interaction. He received the Ph.D. degree in computer science from KAIST in 2013. Contact him at ihwang@us.ibm.com.

Youngki Lee is an assistant professor at Seoul National University. His research interests include innovative software systems over a wide spectrum across systems, applications, and users. He received the Ph.D. degree in computer science from KAIST in 2013. He is the corresponding author. Contact him at youngkilee@snu.ac.kr.

Chungkuk Yoo is a research staff member at IBM. He received the Ph.D. degree from KAIST in 2018. Within the broad spectrum of mobile computing, his research interests include mobile applications for *in situ* social interaction. He is also interested in mobile systems for visual sensing. Contact him at ckyoo-@ibm.com.

Chulhong Min is a research scientist at Nokia Bell Labs, Cambridge, U.K. His research interests include mobile and wearable systems, and the Internet of Things. He received the Ph.D. degree in computer science from KAIST in 2016. Contact him at chulhong.min@nokia-bell-labs.com.

Dongsun Yim is an associate professor of Communication Disorders, Ewha Womans University and a certified Speech-Language Pathologist. Her research interest focuses on the relationship between nature and nurture piece of language learning, as well as providing clinical services to individuals with communication disorders. She received the Ph.D. degree from University of Minnesota. Contact her at sunyim@ewha.ac.kr.

John Kim is an associate professor with the School of Electrical Engineering, KAIST. He received the Ph.D. degree from Stanford University. Previously, he worked on the design of several processors at Motorola and Intel. His research interest includes computer architecture, interconnection networks, and mobile systems. Contact him at jjk12@kaist.edu.