NEMSI: A Multimodal Dialog System for Screening of Neurological or Mental Conditions

David Suendermann-Oeft, Amanda Robinson, Andrew Cornish, Doug Habberstad, David Pautler, Dirk Schnelle-Walka, Franziska Haller, Jackson Liscombe, Michael Neumann, Mike Merrill, Oliver Roesler, Renko Geffarth
d@modality.ai
Modality.AI, Inc.
San Francisco, CA, USA

ABSTRACT

We present NEMSI, a cloud-based multimodal dialog system designed to have naturalistic interactions with individuals for the purpose of screening neurological or mental conditions. The system has been used by thousands of people capturing audio and video responses to open-ended questions and structured health surveys.

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**.

KEYWORDS

dialog systems, multimodal systems, health technology

ACM Reference Format:
David Suendermann-Oeft, Amanda Robinson, Andrew Cornish, Doug Habberstad, David Pautler, Dirk Schnelle-Walka, Franziska Haller, Jackson Liscombe, Michael Neumann, Mike Merrill, Oliver Roesler, Renko Geffarth. 2019. NEMSI: A Multimodal Dialog System for Screening of Neurological or Mental Conditions. In *IVA '19: ACM International Conference on Intelligent Virtual Agents, July 02–05, 2019, Paris, France.* ACM, New York, NY, USA, 3 pages. https://doi.org/?

1 INTRODUCTION

Neurological and mental conditions are a growing concern to society. E.g., the death rate among the middle-aged white population in the U.S. caused by "despair" (suicide or substance abuse) has doubled over the last two decades [2]. An estimated 16 million people in the U.S. are now affected by depression disorder [4]. During a similar interval, death rates from Alzheimer's disease climbed by 55%, now affecting nearly six million Americans [10].

For both aforementioned example conditions it is true that early detection can mitigate the issue. Early treatment of depression lowers the risk of suicide, boosts productivity, and lowers health costs [7]. In the case of Alzheimer's disease, the benefits of early detection include behavioral stabilization, preserved independence, and slowed cognitive decline [5].

However, early detection of neurological or mental conditions is often not possible since many people do not have access to neurologists or psychiatrists where they live, they may not be aware of that they should be seeing a specialist, there is often substantial transportation and cost involved, and there is a severe shortage of medical specialists in these fields to begin with. As a solution, we are presenting NEMSI (NEurological and Mental health Screening Instrument), a cloud-based multimodal dialog system that conducts automated screening interviews over the phone, smart phone app, or web browser to elicit evidence required for detection of the aforementioned conditions, among others.

Intelligent virtual agents have been considered for use for clinical applications, including in areas of mental, behavioural, or neurological health in the past few years. For example, SimSensei Kiosk [3] is a virtual human interviewer specifically built to render clinical decision support. It captures verbal and non-verbal behaviors to extract distress indicators correlated with mental conditions such as depression or PTSD. SimSensei Kiosk deploys real-time computer vision and speech analytics capabilities to control dialog management and non-verbal behavior of the avatar. The study showed that people are generally comfortable sharing personal information with a virtual agent, even if speech and video signals are being captured.

[8] presented results of a large-scale effort, funded by several NSF grants, building a virtual health assistant for "brief motivational interventions", for example, interviews about a subject's drinking behavior. The described system is using text input from the subject's keyboard (or, alternatively,

IVA '19, July 02–05, 2019, Paris, France

ACM ISBN ?...$?
https://doi.org/?

a speech recognition hypothesis) along with the information about the subject's facial expressions to determine what next steps to undertake in the interaction. To produce avatar behavior that is adequate and empathetic, contents of user responses and their facial expressions are interpreted in real-time and influence the avatar's action units.

The NEMSI system draws upon findings from the aforementioned studies, but makes three significant contributions:

- The above systems require dedicated, locally administered hardware, including cameras, servers, audio devices, etc. such that the end user can interact with it. In contrast, the NEMSI system uses end points available to everyone everywhere (web browser, mobile app, or regular phone).
- Furthermore, NEMSI's backend is deployed in an automatically scalable cloud environment allowing it to serve an arbitrary number of end users at a very small cost per interaction.
- Thirdly, the NEMSI system is natively equipped with real-time speech and video analytics modules [9] extracting a variety of features of direct relevance to clinicians in the neurological and mental spaces (such as speech and pause duration, that are markers to assess amyotrophic lateral sclerosis [6], or geometric features derived from facial landmarks for the automated detection of orofacial impairment in stroke [1].)

2 THE NEMSI SYSTEM

As motivated in the introduction, the NEMSI system was designed to deliver screening sessions at scale as conveniently as possible, while providing central access to results and session details to providers for diagnosis or analysis. Consequently, it was decided to move speech and video processing components as well as dialog management into the backend (the cloud) where they can be easily modified and scaled, and where captured data and screening results can be kept on encrypted central storage from which only authorized individuals (such as the responsible physician) can access them.

Figure 1 provides a high-level breakdown of the major components, including

- endpoints (web browsers, native apps, and regular telephony)
- telephony servers handling concurrent traffic, routing, scaling, recording
- a speech server managing voice activity detection, speech recognition, speech synthesis, and playback of prerecorded audio
- real-time speech and video analytics modules extracting features relevant for neurological or mental health screening

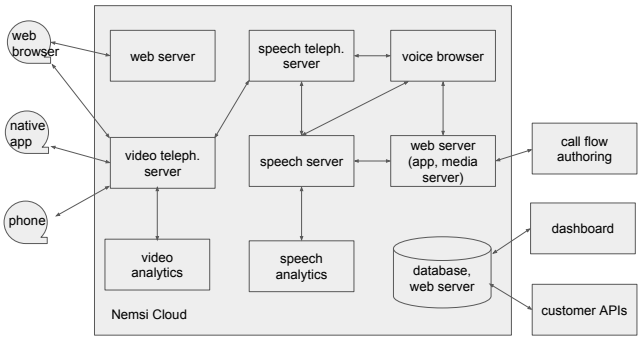

Figure 1: A schematic of the NEMSI system.

- a voice browser (aka dialog manager) interpreting the call flow (i.e. the interaction logic)
- web servers and database to serve the call flow to the voice browser, host model and audio files, and store system logs, meta data, etc.
- a call flow authoring suite that allows to craft the call flow in a straightforward fashion using a mix of graphical and code components
- a dashboard providing clinicians with summary statistics as well as results of the audio-visual analysis carried out during the interaction

3 NEMSI IN ACTION

As indicated in the previous section, NEMSI can engage with end users (patients) through several end points. To illustrate a typical use of NEMSI, we will walk the reader through a scenario in which an end user is using a web brower portal for the screening interaction.

End users are provided with a website link to the secure screening portal as well as login credentials by their provider (physician or clinic). To carry out a session, they log on, and first pass a microphone, speaker, and camera check to assure that the captured signals are of sufficient quality and that users can hear the dialog system's voice. Once passed, they launch the session by clicking the "Start Conversation" button as shown in Figure 2. This enables the camera whose captured video screen is shown back to users in a small window in the upper right corner of the screen. At the same time, the interactive speaking session is started with the avatar introducing herself and setting the stage. She then engages with users in a conversation using a mixture of structured speaking exercises and open-ended questions to elicit speech and facial behaviors relevant for the type of condition being screened for. For example, the avatar may ask users:

> Say /pataka/ ten times as fast as possible

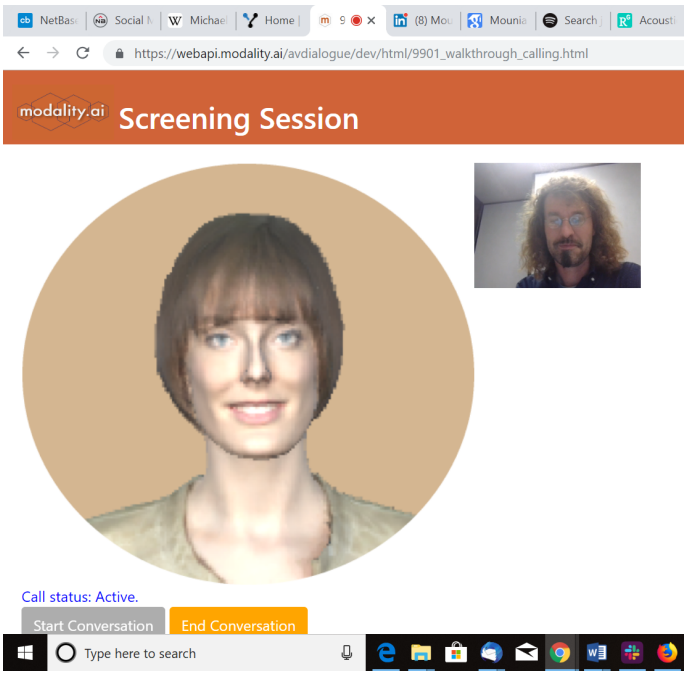

Figure 2: A screenshot of the NEMSI front end during an interactive session.

which is a typical exercise when assessing ALS or other neurological conditions. On the other hand, open-ended questions provide a way to capture extended speech responses of sometimes over one minute duration, allowing to apply a variety of speech and also language measures, e.g.:

> Please describe your current living situation and how you feel about it. Do you live alone or with family or friends?

While the interaction is in progress, the speech and video analytics modules extract the aforementioned features and store them in the database, along with information about the interaction itself such as the captured user responses, the route the system took through the call flow, call duration, completion status, etc. All this information can be accessed by the clinicians after the interaction is completed through a different login to the screening portal. They are able to browse through sessions of their patients, get a high-level overview of analytics results through a dashboard, see the interaction details, and even watch the entire session's video.

4 CONCLUSION

We presented NEMSI, a multimodal dialog system for screening of neurological and mental conditions. The system that is currently under development has already been used by thousands of people and processed over ten thousand interactions, capturing over 500 hours of audio and video material. We are working with several clinical partners to test NEMSI in realistic clinical conditions and run comprehensive studies on the accuracy of its analytics results, as well as its effectiveness and efficiency.

5 ACKNOWLEDGMENTS

We would like to thank our partners Ellipsis Health, the University of Texas Southwestern Medical Center, and Massachusetts General Hospital Institute for Health Professions for their continuous support, as well as the thousands of users that help us improve NEMSI week after week.

REFERENCES

[1] Andrea Bandini, Jordan Green, Brian Richburg, and Yana Yunusova. 2018. Automatic Detection of Orofacial Impairment in Stroke. *Proc. Interspeech 2018* (2018), 1711–1715.

[2] Anne Case and Angus Deaton. 2017. Mortality and morbidity in the 21st century. *Brookings papers on economic activity* 2017 (2017), 397.

[3] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1061–1068.

[4] Statista Dossier. 2017. Depression in the U.S. – Statistics & Facts. Retrieved March 8, 2019.

[5] SG Gauthier. 2005. Alzheimer's disease: the benefits of early treatment. *European Journal of Neurology* 12 (2005), 11–16.

[6] Jordan R Green, Kristen M Allison, Claire Cordella, Brian D Richburg, Gary L Pattee, James D Berry, Eric A Macklin, Erik P Pioro, and Richard A Smith. 2018. Additional evidence for a therapeutic effect of dextromethorphan/quinidine on bulbar motor function in patients with amyotrophic lateral sclerosis: A quantitative speech analysis. *British journal of clinical pharmacology* 84, 12 (2018), 2849–2856.

[7] Aron Halfin. 2007. Depression: the benefits of early and appropriate treatment. *The American journal of managed care* 13, 4 Suppl (2007), S92–7.

[8] Christine Lisetti, Reza Amini, and Ugan Yasavur. 2015. Now all together: overview of virtual health assistants emulating face-to-face health interview experience. *KI-Künstliche Intelligenz* 29, 2 (2015), 161–172.

[9] Oliver Roesler and David Suendermann-Oeft. 2019. Towards visual behavior detection in human-machine conversations. In *Proceedings of the 2019 international conference on Activity and behavior computing*.

[10] Christopher A Taylor, Sujay F Greenlund, Lisa C McGuire, Hua Lu, and Janet B Croft. 2017. Deaths from Alzheimer's Disease–United States, 1999–2014. *MMWR. Morbidity and mortality weekly report* 66, 20 (2017), 521.