

No-code Digital Human for Conversational Behavior

Hanseob Kim
Korea Institute of Science
and Technology
Seoul, Republic of Korea
khseob0715@kist.re.kr

Jieun Kim
Korea Institute of Science
and Technology
Seoul, Republic of Korea
092296@kist.re.kr

Ghazanfar Ali
University of Science and
Technology
Seoul, Republic of Korea
aliust@ust.ac.kr

Jae-In Hwang
Korea Institute of Science
and Technology
Seoul, Republic of Korea
hji@kist.re.kr

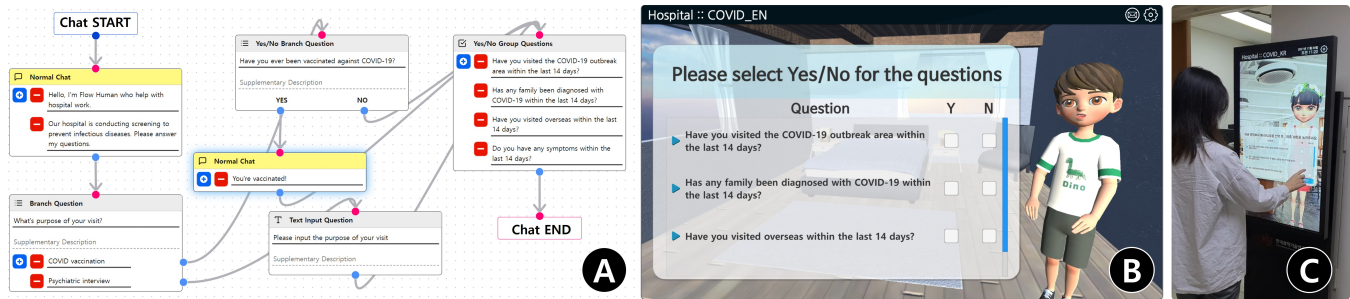


Figure 1: Flow Human system overview: (a) Flow-based dialogue authoring tool; (b) Digital human with conversational behavior and feedback form UI; and (c) The photograph of a user interacting with Flow Human embedded in the Kiosk (as an use case).

ABSTRACT

In this poster, we present Flow Human, a no-code system that generates conversational behavior of digital humans from the text. Our users only need to build a conversation flow they want to talk to customers using the flow-based authoring tool we developed. Our system then automatically generates the verbal and non-verbal behavior of digital humans along the conversation flow, interacts with customers, and collects feedback. We believe that this work can serve the potential to be distributed to various services that have not been introduced because of the challenging task of controlling multiple factors in digital humans (e.g., conversation flow, co-speech gestures, and facial animation).

CCS CONCEPTS

• **Human-centered computing** → **User interface toolkits**; *Personal digital assistants*.

KEYWORDS

Flow Human; No-code System; Digital Human; Flow-based Toolkit.

ACM Reference Format:

Hanseob Kim, Jieun Kim, Ghazanfar Ali, and Jae-In Hwang. 2022. No-code Digital Human for Conversational Behavior. In *SIGGRAPH Asia 2022 Posters (SA '22 Posters)*, December 06-09, 2022, ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3550082.3564175>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SA '22 Posters, December 06-09, 2022, Daegu, Republic of Korea

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9462-8/22/12.

<https://doi.org/10.1145/3550082.3564175>

1 INTRODUCTION

Today, people expect digital humans to behave like humans because they have a human-like appearance [Kim et al. 2021]. Therefore, there is a desire in the industry to replace real people by allowing digital humans to play a social role in repetitive tasks or hazardous sites (e.g., Infectious disease screening explainer). However, developing digital humans that allow users to engage in social interactions is challenging. This is because there are many factors to consider in order to imitate real human capabilities [Harttholt et al. 2013]. The industry, for example, may be able to play digital humans' voice with sound devices but may not be prepared for the conversational behavior (e.g., mouth movements, hand gestures) that accompanies the voice, i.e., software-side factors. In this case, people may lose trust in digital humans and not have meaningful interactions.

We present Flow Human, a no-code system that automatically generates digital humans' conversational behavior. We developed a flow-based authoring tool to allow users to build a conversation flow to be spoken by digital humans. Our system uses dialogue phrases to generate mouth movements and hand gestures accompanying the voice. Thus, our users can provide compelling digital humans to their customers by building only the conversation flow. We also provide a customer feedback collector to enable users to apply our digital human to various contexts (see Section 3). The main novelty is that we are the first to present a no-code digital human system for conversational behavior. The goal of this work is to improve the difficulty of introducing digital humans in a user-friendly way.

2 IMPLEMENTATION

Flow Human system consists of three parts: (1) Flow-based dialogue authoring tool; (2) Conversational behavior of digital human; and lastly, (3) Collecting customer feedback. Our Flow Human is a lightweight system that is deployed as a web page. We provide a demonstration video (<https://youtu.be/z9camFP59YE>).

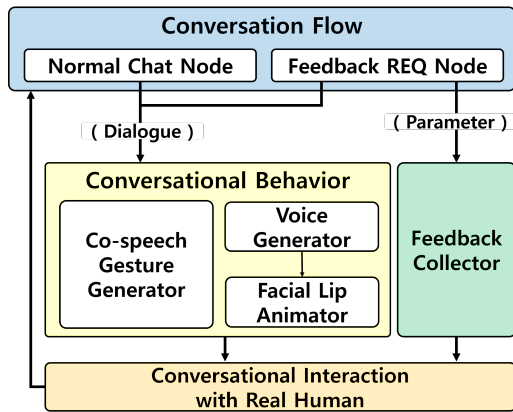


Figure 2: Architecture of the Flow Human System.

2.1 Flow-based Dialogue Authoring Tool

Our system requires users to do only two tasks: (1) selecting a digital human and (2) creating a conversation flow for their customers. To begin with, users create a digital human to interact with their customers. Users can choose the embodiment and voice of digital humans according to their service type or preferences from our library. Our digital humans have dynamic behaviors (e.g., eye-blinking, head movements, and idle behaviors) that are not bound by the conversation flow. The selected digital human is then displayed alongside the flow editing tool on the same page.

Users then build a conversation flow using a flow-based authoring tool (see Figure 1 (a)). A conversation flow is a set of visually connected dialogue nodes. These nodes are divided into two types. The first is a normal chat node where users can write a dialogue list. The second is the feedback request node, where users can write dialogue and parameters for collecting customer feedback (e.g., typing a message, selecting an item, and weighted value). Our users can drag and drop these nodes onto the canvas and connect their two edges to build a conversation flow to be spoken by digital humans.

2.2 Conversational Behavior of Digital Human

In this work, we generate lip motions and co-speech gestures that accompany the voice among the conversational behavior of digital humans. Our system uses the dialogue phrase, which is text data, typed by the user passed from the conversation flow. First of all, for voice generators, we used a text-to-speech API¹. Our system then uses the generated voice file to animate phoneme-based lip motion with 7 mouth forms and intensities ranging from 0 to 100.

For the co-speech gesture, we modified text-to-gesture rulemap method [Ali et al. 2020]. This method first obtain (text, pose) pairs from public monologue videos using Speech-To-Text and pose estimation. In second step, pose sequences are chunked and mapped to predefined set of gesture animations. Reason being pose sequence from videos are usually noisy hence unusable directly. We added total of 2035 new gestures with 1025 for male and 1010 for female style. Sentence-BERT [Reimers and Gurevych 2019] model is used to improve the contextual and semantic text matching at runtime.

¹<https://www.ncloud.com/product/aiService/cpv>

2.3 Collecting Customer Feedback

Our system provides various types of feedback forms to customers when the current node of the conversation flow is the feedback request node (see Figure 1 (b)). Customer feedback can be used to control the conversation flow using the selected items or weighted-value during the interaction, or for post-analysis.

3 USE CASE

3.1 Survey and Screening Process

Flow Human can conduct in-depth interviews and elicit active responses. Our conversation flow can be branched based on the respondent's answers and/or weighted values, making it easier to ask additional questions and provide quick feedback. Furthermore, thanks to various and appropriate feedback by branched sub-flows and weighted values, our system can induce the active participation of respondents tired of repetitive tasks.

3.2 FAQs and Support

Digital humans can replace customer service personal. Users can build a conversation flow by organizing frequently asked questions from their customers. Thus, if a customer interacts with a digital human, they can request feedback without waiting time.

3.3 Product and Service Explainer

Users can build a conversation flow that introduces products and deploy digital humans in their stores or websites. Then customers can always hear the product introduction and only listen to what they are questions about the product. Furthermore, our system would be helpful to find out what features their customers were interested in without the influence of employees' gaze.

4 CONCLUSION

We present Flow Human, a no-code digital human system that can generate conversational behavior. Our users only need to build the conversation flow using our authoring tools. Users can easily prepare and provide a compelling digital human to serve their customers. We plan to explore usability through formal user studies.

ACKNOWLEDGMENTS

This work was supported by the National Research Council of Science & Technology grant by the Korea government (NO. CRC-20-02-KIST) and also by the Korea Creative Content Agency in the Culture Technology Research & Development Program.

REFERENCES

- Ghazanfar Ali, Myungho Lee, and Jae-In Hwang. 2020. Automatic text-to-gesture rule generation for embodied conversational agents. *Computer Animation and Virtual Worlds* 31, 4-5 (2020), e1944. <https://doi.org/10.1002/cav.1944>
- Arno Hartholt, David Traum, Stacy C Marsella, Ari Shapiro, Giota Stratou, Anton Leuski, Louis-Philippe Morency, and Jonathan Gratch. 2013. All together now. In *International Workshop on Intelligent Virtual Agents*. Springer, 368–381.
- Hanseob Kim, Myungho Lee, Gerard J. Kim, and Jae-In Hwang. 2021. The Impacts of Visual Effects on User Perception With a Virtual Human in Augmented Reality Conflict Situations. *IEEE Access* 9 (2021), 35300–35312. <https://doi.org/10.1109/ACCESS.2021.3062037>
- Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. <https://arxiv.org/abs/1908.10084>