Musical and Conversational Artificial Intelligence

Fabio Catania
fabio.catania@polimi.it
Politecnico di Milano

Giorgio De Luca
giorgio.deluca@mail.polimi.it
Politecnico di Milano

Nicola Bombaci
nicola.bombaci@mail.polimi.it
Politecnico di Milano

Erica Colombo
erica1.colombo@mail.polimi.it
Politecnico di Milano

Pietro Crovari
pietro.crovari@polimi.it
Politecnico di Milano

Eleonora Beccaluva
eleonora.beccaluva@polimi.it
Politecnico di Milano

Franca Garzotto
franca.garzotto@polimi.it
Politecnico di Milano

ABSTRACT

Music production software often has complex interfaces and needs the user to know the basic musical know-how. In this paper, we present a conversational agent that allows creating music in a simplified way through voice-based interaction. Indeed, our agent can be configured and customized with simple and natural voice commands. In addition, it has some typically human cognitive skills to produce music: it listens to the user while singing a song and generates a melody by discovering and copying the patterns of her/his human voice. Technologically, the system is empowered by Google Dialogflow for conversation management and uses an advanced technique called abstract melody for music production. This Musical and Conversational Artificial Intelligence is an actual innovation since it does not require any preliminary knowledge about music and, consequently, includes professionals, but also children, beginners, and people with physical disease.

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**; • **Applied computing → Sound and music computing**.

KEYWORDS

Conversational technology, Conversational interface, Sound and music computing, Human Computer Interaction

ACM Reference Format:
Fabio Catania, Giorgio De Luca, Nicola Bombaci, Erica Colombo, Pietro Crovari, Eleonora Beccaluva, and Franca Garzotto. 2020. Musical and Conversational Artificial Intelligence. In *25th International Conference on Intelligent User Interfaces Companion (IUI '20 Companion), March 17–20, 2020, Cagliari, Italy*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3379336.3381479

1 INTRODUCTION

Musical intelligence is one of Gardner's nine different kinds of intelligence that characterize human beings. [3] It is dedicated to how skillful an individual is at performing, composing, and appreciating music and musical patterns. Not surprisingly, musicians, composers, band directors, disc jockeys and music critics are among those that Gardner sees as having high musical intelligence. In recent years, technological progress made it possible for these people and others to produce music with the help of technology. [5, 7] However, nowadays, people who want to make good and original music using technology must know the basic principles of music and are required to work with music software, that often has complex interfaces [1].

In this paper, we present a conversational agent that allows creating music through an easy-to-use voice-based interface. By conversational agent (CA), we mean a dialogue system able to interact with a human through natural language [2]. Our agent is capable of

- asking questions about the music the user wants to produce in a non-technical way,
- listening to the user while singing a song and analyzing her/his pitch,
- and finally generating a tuned melody by using the information obtained during the conversation and the musical patterns discovered from her/his singing voice.

Technologically, our Musical and Conversational Artificial Intelligence can imitate the typically human cognitive skills to produce music by using an advanced technique called abstract melody [4].

Our product may have many potentials. Technavio states that the emergence of AI in music composition is expected to have a positive impact on the growth of the global music production market during the forecast period. [9] A conversational interface like ours can be used to produce music

- by children and beginners who want to approach the world of music,
- by professionals with a jingle in mind,
- by people with motor disabilities who cannot use musical instruments or other technological interfaces that are not accessible to them.

2 CONTEXT

Nowadays, there are different software for professional music production: Digital Audio Workstation or DAW (e.g., *DAW like Logic Pro*, *ProTools*, *Reaper*), programs to process or manipulate audio

IUI '20 Companion, March 17–20, 2020, Cagliari, Italy

ACM ISBN 978-1-4503-7513-9/20/03.
https://doi.org/10.1145/3379336.3381479

with effects (e.g., *Audition*, *Audacity*), and mobile apps like drum machines or simple editors with pre-built set of musical parts (e.g., *SNAP*, *Patterning 2*, *Impaktor*). Most of these applications are quite difficult to be used [1], since they provide long lists of functions, different types of view perspective (time, beats, samples), routing systems, and chains of effects. In some sense, this technology complexity may represent an entrance barrier for new users.

Speaking about amateur music composition, a huge amount of simple applications are available (e.g., *Steinberg Cubasis*). They provide limited, seemingly nonsensical functions, such as the possibility to combine predefined patterns and tuned melodies to create a composition. Few programs for beginners give space to user creativity; they usually have interfaces similar to a professional DAW, but with limitations. For example, *GarageBand* allows the user to personalize the melody through a piano roll (a representation of notes in time).

As far as we know, the only existing software on the market able to record your voice and add a genre-oriented accompaniment is *HumOn*. It is a mobile application, it can be used with the touch, and it is based on the harmonization of a melody provided by the user: this means that, given a melody hummed by the user, the system creates a series of chords that fit and sound good with the melody, which is not modified.

Entering into the technical merits, we could not find any program able to take a melody and a chord progression and to create a melody that fits the chord progression by choosing the notes from the scale defined by the chords. In addition, to the best of our knowledge, there is no software for music production managed with a natural and intuitive interface mostly based on the use of voice. Our study aims to investigate these last two unexplored fields.

3 CONVERSATIONAL INTELLIGENCE

The user can interact with our musical intelligence through a natural and intuitive conversational interface. In taxonomy, our dialogue system is goal-oriented, domain restricted, and proactive. The conversational flow is standardized, since our agent asks some specific questions to the user. In particular, she/he is invited to choose the *time*, the *chords*, the *progression*, and the *scale mode* of the piece she/he wants to compose. To make the application accessible to people with little/no musical background, technical concepts (if needed) are accompanied by simple explanations and questions are formulated as a set of options where the user has to respond with the preferred choice. Besides asking questions, the agent lets the user hum a melody, which is the starting point for the piece to be produced. When the song has been generated, the user can request the system to play it, to edit it and to combine several tracks.

The prototype of the system has been realized as a web application. Speeches by the user are recorded on the client and then are sent to the server to be processed. Speech-to-Text, Natural Language Understanding, and dialogue management are performed by exploiting Dialogflow by Google. The agent's vocal answers are generated on the client by exploiting ReadSpeaker.js, a text-to-speech javascript library.

4 MUSICAL INTELLIGENCE

Our musical intelligence uses the Short-Time Fourier Transform to analyze the voice by the user while singing a song. As a result, it generates a MIDI file, which is a standard instructional file that illustrates which notes are played, when they are played, and how long and loud each note is. Now, the system uses an advanced technique called abstract melody [4] to extract the progress of the pitch in the time domain staring from the specifications from the MIDI file. The pitch includes information about its variation in time and the presence/absence of a note in every single time instant. At this point, the system obtains the *reference scale* for the final melody directly from the chord selected by the user during the conversation. This is possible because a chord is just a sub-sample of a scale [6, 8]. Finally, the musical intelligence follows an original method to arrange the notes in the MIDI file so that the final song sounds in tune with the reference scale. The tone of the first note of the output melody is chosen randomly from the notes of the selected scale in the previous phase. The tone of the next notes is chosen among the notes included within the distance (in terms of semitones) between the last note in the melody generated so far and the note under analysis in the MIDI file. Specifically, the music intelligence associates strong grades to long notes and weak ones to short notes, in order to create a melody containing rapid tension changes followed by an optimal resolution to the strong grades of the scale.

5 CONCLUSION

In this paper, we faced the difficulty in producing music with the help of technology without using predefined patterns, especially for people with little/no musical knowledge and who are not familiar with the use of music software, which often has complicated interfaces. We developed a conversational agent with musical intelligence that can take voice-based commands, analyze melodies sung by the user, and turn everything into a beautiful and harmonically corrected song. Technologically, our musical Artificial Intelligence uses an advanced technique called abstract melody.

In the next future, we plan to test the usability of our technology with different populations: professional adults, non-professional adults, children, and disabled people. Besides, we would like to compare the level of user satisfaction, the grade of engagement, and the quality of the musical productions made with our musical intelligence and those made with similar technologies.

REFERENCES

[1] Richard James Burgess. 2013. *The art of music production: the theory and practice.* Oxford University Press.
[2] DeepAI. 2019. Conversational Agent. Online, www.deepai.org/machine-learning-glossary-and-terms/conversational-agent.
[3] Howard Gardner. 1987. The theory of multiple intelligences. *Annals of dyslexia* (1987), 19–35.
[4] Jon Gillick, Kevin Tang, and Robert M. Keller. 2010. Machine Learning of Jazz Grammars. *Comput. Music J.* 34, 3 (Sept. 2010), 56–66.
[5] Peter Knees, Markus Schedl, and Rebecca Fiebrink. 2019. Intelligent music interfaces for listening and creation. In *Proceedings of the 24th International Conference on Intelligent User Interfaces: Companion.* ACM, 135–136.
[6] Mark Levine. 2011. *The jazz theory book.* " O'Reilly Media, Inc.".
[7] Anupam Shukla, Ritu Tiwari, and Rahul Kala. 2010. Intelligent Systems Design in Music. In *Towards Hybrid and Adaptive Computing.* Springer, 153–173.
[8] Peter Spitzer. 2015. *Jazz theory handbook.* Mel Bay Publications.
[9] technavio. 2019. *Music Production Software Market by Type, End-users, and Geography - Forecast and Analysis 2020-2024.* technavio.