Boris: a Spoken Conversational Agent for Music Production for People with Motor Disabilities

Fabio Catania
Politecnico di Milano
Milano, MI, Italy
fabio.catania@polimi.it

Pietro Crovari
Politecnico di Milano
Milano, MI, Italy
pietro.crovari@polimi.it

Eleonora Beccaluva
Politecnico di Milano
Milano, MI, Italy
eleonora.beccaluva@polimi.it

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

Erica Colombo
Politecnico di Milano
Milano, MI, Italy
erica.colombo@mail.polimi.it

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

Franca Garzotto
Politecnico di Milano
Milano, MI, Italy
franca.garzotto@polimi.it

ABSTRACT

Previous studies suggest that engagement in musical activities may enhance well-being and impact social inclusion. However, unfortunately, people with physical disabilities cannot often use musical instruments or music production software due to accessibility issues. We propose Boris, an original conversational agent specific for people with a physical disability, to entertain, stimulate expressiveness, and promote communication. Boris enables (even inexperienced) users to compose songs through hands-free interaction by analyzing their vocalizations to obtain more than just their transcription: the system listens to the user even while humming a song and generates a melody by learning and reproducing their human voice patterns. Indeed, it exploits an artificial musical intelligence that can imitate the typically human cognitive skills to produce music using an advanced technique called abstract melody.

CCS CONCEPTS

• **Human-centered computing → Natural language interfaces**;
• **Social and professional topics → People with disabilities**.

KEYWORDS

conversational agents, accessibility, inclusion, music production

ACM Reference Format:
Fabio Catania, Pietro Crovari, Eleonora Beccaluva, Giorgio De Luca, Erica Colombo, Nicola Bombaci, and Franca Garzotto. 2021. Boris: a Spoken Conversational Agent for Music Production for People with Motor Disabilities. In *CHItaly 2021: 14th Biannual Conference of the Italian SIGCHI Chapter (CHItaly '21), July 11–13, 2021, Bolzano, Italy*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3464385.3464713

CHItaly '21, July 11–13, 2021, Bolzano, Italy

ACM ISBN 978-1-4503-8977-8/21/06…$15.00
https://doi.org/10.1145/3464385.3464713

1 INTRODUCTION

Motor disability is any physical condition that makes it more difficult for a person to do certain activities or interact with the external world [24]. It includes muscle weakness, poor stamina, lack of muscle control, or total paralysis. Motor disability often manifests in neurological conditions such a cerebral palsy, Parkinson's disease, stroke, and multiple sclerosis [24]. To give some numbers, 1% of the world population have severe motor disabilities and need a wheelchair on a daily basis. If we count mild and medium forms of disabilities, the percentage is dramatically higher [24].

There are many previous studies about technologies to specifically assist and help people with motor disabilities in the context of everyday life [1, 8] and education: technology not only targets teaching related to a certain content area but also might focus on limiting the difficulties caused by a disability [15]. Technologies are of different natures and involve, among others, conversational technologies [28], virtual reality tools [23], and tangible smart objects [26]. In contrast to assistance and education, there are few tools for fun and entertainment accessible to people with a motor disability so far. This lack was highlighted even by the European Google Challenge, which was organized in collaboration with the NEMO Clinical Center of Milan, Italy, in December 2019 [13]. The competition's goal was to make up for this lack and develop conversational technologies that offer fun opportunities to people with neuromuscular disorders and motor disabilities. The shortage of opportunities for fun and entertainment for people with disabilities is paradoxical if we consider that people of this group spend more time on leisure activities than their non-disabled counterparts [25]. Among all, musical activities are encouraged since evidence suggests that engagement in this discipline may enhance well-being and impact social inclusion [18, 31].

In this context, we propose Boris, a conversational agent that can hold small talks with the user and help her/him in music production. By conversational agent (CA), we mean a dialogue system able to interact with a human through spoken natural language [7]. A conversational interface's strength is that it is accessible to all those who are verbal but with motor impairments. Besides, from literature, we know that speech impairments are widespread among people with some physical disabilities [17]; in response, Boris enables the interaction even by users with severe speech impairments by analyzing their vocalizations to obtain more than just the transcription: the system exploits a novel musical artificial intelligence that can imitate the typically human cognitive skills to produce music by using an advanced technique called abstract melody [4]. Consequently, the user can sing or hum a melody with her/his voice, and the system uses it to compose a song that sounds correct from a harmonic point of view. No singing skills or music knowledge is needed.

To the best of our knowledge, Boris is the first application for music production with a voice-based interface designed to be accessible to people with physical disabilities. Its goal is to stimulate expressiveness, bring people closer to music, entertain, and promote inclusion. This exploratory project has been realized through a user-centered design process with human-computer interaction experts, psychologists, linguistics experts, and a person with severe mobility challenges.

In this paper, we describe Boris's design process, user experience, and software architecture. In the end, we introduce some interesting future works with Boris. This work is far from being mature or extensively technically validated but wants to pave the ground for conversational agents for people with disabilities that explore more than just semantic analysis of speech.

2 STATE OF THE ART

Nowadays, laptop applications panorama is full of tools for creating and producing music. We analyzed the tools on the market and divided them into two categories: professional and amateur technologies.

Professional software provides long lists of functions, different types of view perspective (time, beats, samples), routing systems, and chains of effects. They are Digital Audio Workstation or DAW (e.g., *DAW like Logic Pro*, *ProTools*, *Reaper*), programs to process or manipulate audio 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*). Amateur applications provide just limited simple functions as the possibility to combine predefined patterns and tuned melodies to create a composition (e.g., *Steinberg Cubasis*). For example, *GarageBand* allows the user to personalize the song through a piano roll (a representation of notes in time). *HumOn*, instead, is a mobile application (whose navigation is based on the touch paradigm) and records the user's voice while singing and adds a genre-oriented accompaniment to it. In other words, 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. All the listed professional and amateur software exploit physical interaction on a graphical user interface with the mouse or a touch screen. Consequently, they require the ability to point, press, and drag items. This aspect makes these tools generally (too) hard to be used in autonomy by disabled people [24].

For what concerns open-source web technologies, Chrome Music Lab [12] is an application that makes learning music more intuitive and fun with hands-on experiments. Many teachers use it in their classrooms to explore music and its connections to science, math, art, and more [12]. Again, however, interaction always requires physicality.

The field of musical tools for people with physical disabilities is still a growing emerging area of research [10, 14, 21]. Anderson and Smith [2] studied the possibility of adapting existing tools to be accessible to people with visual and physical impairments. Larsen et al. [20], Parke-Wolfe et al. [27] built a software toolkit that enables music therapists and teachers to create custom digital sensor- and vision-based musical interfaces for children with diverse disabilities. EyeMusic [16] is both a performance and a playback instrument. It uses an eye-tracker that outputs the gaze position (x, y) 60 times per second and operates with eye fixation, which has two parameters: deviation and duration. Brainfingers [9] is a hands-free computer control for music creation: a headband fitted with sensors detects electrical signals from facial muscles, eye movement, and brain waves. To the best of our knowledge, there is no software for music production that is addressed to people with motor disabilities and exploits the conversational channel.

3 THE DESIGN PROCESS

Boris was explicitly designed for people with motor disabilities. As an exploratory work, to understand the difficulties of these subjects, we started by analyzing the needs of Andrea (name of fantasy to guarantee his privacy), who is a 45-year-old man with spastic tetraplegia and language difficulties. Spastic tetraplegia is a subset of spastic cerebral palsy that affects all four limbs (both arms and legs) [19]. We designed Boris through a user-centered process involving human-computer interaction experts, psychologists, linguistics experts, and Andrea. An iterative, agile, three-stage (exploration, implementation, and testing) research process was used. Exploration refers to the detection of the user's preferences and needs to be satisfied. For this reason, we reviewed the literature on the field and held four two-hour meetings with experts and two one-hour sessions with Andrea and his parents. We found out that spastic tetraplegia limits Andrea in many aspects of everyday life. For example, he needs to be helped during lunch to drink and eat or when he wants to send an e-mail with the PC. His reading and writing skills are discrete. Andrea's verbalization is relatively slow and hard, and he often needs to repeat the same sentence several times to be fully understood. These phonological and articulatory impairments are quite common in degenerative pathologies [6] [29], and affect not only Andrea but also many other subjects with different conditions [30]. The implementation phase consisted of the realization of a prototype. Testing was to verify the accessibility of the prototype and its usability by Andrea, and it was also helpful to identify relevant modifications to be implemented in the future. So far, we conducted just one iteration of the process, and, at this point, we observed that Andrea can use the application in autonomy, despite his verbal difficulties. Unfortunately, we could not run

an additional empirical study with a larger population because of the ongoing pandemic, but we plan to do it as soon as possible.

4 THE CONVERSATIONAL EXPERIENCE

Boris's prototype is a web application because of web apps' pervasiveness, ease of use, and the absence of installation and configuration. Since Boris is a web app, it enables both vocal and visual interactions through the screen, microphone, and speakers of the device (both standalone and mobile). Indeed, the user interacts with the system with the only use of the voice, but our conversational agent exploits the visual channel as a communication support. The graphical theme of the application is inspired by space because Boris plays the role of an astronaut who guides the user in the discovery of the *universe* of music. A screenshot of the interface is shown in Fig.1. In taxonomy, Boris is goal-oriented, domain restricted, and proactive. Boris was designed to be as entertaining and engaging as possible. With this goal in mind we decided to designed Boris with a funny personality: it makes jokes and speaks with a marked foreigner accent, but both the syntax and the semantics of the sentences are grammatically correct. According to [22], voice plays an essential role in the perception of conversational agents and impacts the whole user experience. In agreement with Andrea, our choice could be user-activated and useful for "breaking the ice" with the user and create a safe and playful setting To convey this expressiveness in the voice, an actor registered all Boris' utterances. A psychologist and a linguistics expert wrote all dialogues that Boris can hold. Language choices are crucial in dealing with Andrea and with people with disabilities in general since sometimes they experience moments of frustration due to their impairments and have a delicate psychological condition [20]. For that reason, Boris avoids all words related to negative emotional states to provide the user with the safest environment possible.

When the application is launched, our agent introduces itself as an astronaut with a passion for music and explains how it will help the user compose a short song. The agent speaks clearly and with many repetitions. Explaining the same concept many times and in many different ways enables the user to understand the agent and the context better. Boris stimulates and prompts the user with a series of simple questions about the music to compose that can be answered with "yes" or "no" to avoid any misunderstanding due to Andrea's speech impairments. Therefore, the user can feel free from any possible pressure not to complete the task and can use just her/his creativity and musical taste to deal with the choices. Since the system was tuned on Andrea's lack of specific competences in the field of music production, it suits for early stage approach with the music: no technical knowledge or special musical skills are required to reply to any of the questions. Questions on musical topics are presented through practical examples. First, Boris asks the user if she/he prefers a slow or fast tempo. Each option is exemplified by a sample of a metronome beating at two different speeds. The next question is about the song's mood to be composed: "Would you like to make your song sound happy?" or "Would you like to make your song sound sad?". At this point, Boris makes the user listen to a chord progression and asks which one she/he likes best. This question is formulated as a set of examples where the user has to respond with the preferred choice. Finally, Boris

Figure 1: A screenshot of Boris interface. The conversational agent is represented by the astronaut floating around the space. The line on top left corner indicates the progresses in the activity. The box in the bottom-right corner contains the transcription of Boris' utterances, whereas the top-left corner an icon appears when the microphone is active.

Figure 2: Musical Intelligence Pipeline

invites the user to hum a melody used as the starting point for the output song. At this point, the AI generates a melody by learning and reproducing the human voice patterns and using the info about the tempo and the wished song's mood from the previous replies. Once the final song has been produced, the user can request Boris to play it, edit it, and combine several tracks. Combining several songs into one allows users to produce a song alone or in a group.

5 SOFTWARE ARCHITECTURE

Technically speaking, Boris is a web application, and the user's speeches are recorded within the browser and then are sent to the server to be processed. The architecture follows the guidelines by Catania et al. [5]. Speech-to-Text, Natural Language Understanding, and dialogue management are performed by exploiting Dialogflow by Google. The user's vocal answers are the audio recordings by the actor played within the user's browser. The software module for music production is proprietary and is better described as follows.

5.1 The musical intelligence

Boris creates a melody that sounds correct from a harmonic point of view through an artificial emotional intelligence that exploits an advanced technique called *abstract melody* [4]. To do that, it combines the information obtained from both the conversation with the user and the analysis of her/his pitch while singing a song. Following, we go through the whole music production process, summarized in Fig. 2.

First, the system analyses the user's voice by tracking her/his pitch using the Short-Time Fourier Transform (STFT): the procedure is to divide the whole signal into shorter windows of equal duration and then compute the Fourier transform separately on each shorter segment. Once the system has the Fourier spectrum on each window, it finds for each of them the fundamental frequency, which is the lowest harmonic in the spectrum and is the one that

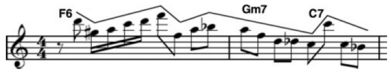

Figure 3: The progress of the pitch in time represented by the broken line above the pentagram.

represents most the pitch. In this way, the changing pitch can be analyzed as a function of time. The pitch includes information about its variation in time and the presence/absence of note in every single time instant. As a result of this phase, the system generates a MIDI file, a standard instructional file that illustrates which notes are played, when they are played, and how long and loud each note is. The system can finally use the abstract melody [11] to extract from the MIDI file the specifications about the distances between consecutive notes. Boris' melody can be imagined as a broken line, going up and down as the input voice goes up or down, like in Figure 3. The distance of the notes is measured in terms of semitones. At this point, the system defines which notes to include in the final melody. This sub-domain of notes is obtained from a scale, called the *reference scale*, that defines both melodic and harmonic aspects in a song. In music, a single chord provides a set of possible equivalent scales, but a set of different chords can define just a single scale. Consequently, the system obtains the reference scale for the final melody directly from the user's chord progression during the conversation. 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 previous phase's selected scale notes. The next notes' tone 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. In music, given a specific scale, the "strong" grades are 1st, 3rd, 5th, and 7th, since they are responsible for the scale identification (for example, a scale in C major is composed by C, D, E, F, G, A, B and is identified by the chord C, E, G, B). The "weak" grades instead are 2nd, 4th, and 6th, and they are used to create tension to lead to an immediate resolution to a strong(er) grade. That said, the system chooses every note from the set of eligible notes by following this policy: it associates strong grades to long notes and weak ones to short notes to create a melody containing rapid tension changes followed by an optimal resolution to the strong grades of the scale.

6 CONCLUSION AND FUTURE WORKS

We presented Boris, a conversational agent for entertainment capable of producing a tuned song from a user's melody thanks to an original artificial musical intelligence. Boris's strengthen is that it responds not only according to the semantic content of the speech but also to the pitch analysis result, as already seen in a previous study with children with neurodevelopmental disorders [3]. Boris was explicitly designed for people with motor disabilities, as they cannot generally use musical instruments or physical musical interfaces due to accessibility issues.

This project is still an early bird, but since our first exploratory and qualitative user-testing with a person with spastic tetraplegia resulted in positively affecting the engagement and facilitating music production, we will surely continue with this work. From our observations, Andrea was happy to play with Boris and succeeded in completing the experience in autonomy answering all Boris' questions, respecting turn-taking times, and humming a short song. Besides, once he had created his song, he invited his parents to join him to introduce new melodies in his creation.

The main limitation of the work consists in the followed *design for one' process*, focused on the needs of a specific individual users. While this approach allows to examine various aspects of the system in depth, it does not guarantee its overall accessibility and usability, also taking into account that people with motor disabilities have very wide range of abilities and functional limitations and there may be difficulties in the processing of their vocalizations due to voice characteristics related to their motor impairment.

As a consequence, the natural follow-up of this work will be a long-term experimentation with target users: we want to verify our application's usability by more people with motor disabilities since it has not been possible due to the ongoing pandemic. It would be interesting to assess user satisfaction, the fatigue in the interaction, the grade of engagement, and the quality of the musical compositions made with our conversational interface. Also, we would like to compare these aspects concerning our tool and other music production technologies, so as to illustrate the extent to which Boris supports inclusive music production with respect to able-bodied individuals. Besides, the application can be used alone or in groups. Consequently, we would like to investigate the ability to promote inclusion by technology like the one we have described in the paper. Finally, we consider investigating whether the Boris approach could be extended to other user categories (e.g., very young children, blind children, older people, etc.).

REFERENCES

[1] Sandra Alper and Sahoby Raharinirina. 2006. Assistive Technology for Individuals with Disabilities: A Review and Synthesis of the Literature. *Journal of Special Education Technology* 21, 2 (2006), 47–64.
[2] Tim Anderson and Clare Smith. 1996. "Composability" widening participation in music making for people with disabilities via music software and controller solutions. In *Proceedings of the second annual ACM conference on Assistive technologies*. 110–116.
[3] Fabio Catania, Nicola Di Nardo, Franca Garzotto, and Daniele Occhiuto. 2019. Emoty: an emotionally sensitive conversational agent for people with neurodevelopmental disorders. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.
[4] Fabio Catania, Giorgio De Luca, Nicola Bombaci, Erica Colombo, Pietro Crovari, Eleonora Beccaluva, and Franca Garzotto. 2020. Musical and Conversational Artificial Intelligence. In *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion*. 51–52.
[5] Fabio Catania, Micol Spitale, Davide Fisicaro, and Franca Garzotto. 2019. CORK: A COnversational agent framewoRK exploiting both rational and emotional intelligence.. In *IUI Workshops*.
[6] Karen Croot, John R Hodges, John Xuereb, and Karalyn Patterson. 2000. Phonological and articulatory impairment in Alzheimer's disease: a case series. *Brain and language* 75, 2 (2000), 277–309.
[7] DeepAI. 2019. Conversational Agent. Online, www.deepai.org/machine-learning-glossary-and-terms/conversational-agent.
[8] Dave L Edyburn. 2000. Assistive technology and mild disabilities. *Mental retardation* 612 (2000), 10–6.
[9] Brain fingers. 2019. Brain fingers. Online, http://www.brainfingers.com/.
[10] Emma Frid. 2019. Accessible digital musical instruments—a review of musical interfaces in inclusive music practice. *Multimodal Technologies and Interaction* 3, 3 (2019), 57.
[11] Jon Gillick, Kevin Tang, and Robert M. Keller. 2010. Machine Learning of Jazz Grammars. *Comput. Music J.* 34, 3 (Sept. 2010), 56–66.

[12] Google. 2019. CHROME MUSIC LAB. Online, https://musiclab.chromeexperiments.com/.
[13] Google. 2019. Google Assistant for Good - European Challenge. Online. shorturl.at/bhvOQ.
[14] Jacob Harrison and Andrew P McPherson. 2017. An adapted bass guitar for one-handed playing.. In *NIME*. 507–508.
[15] Ted S Hasselbring and Candyce H Williams Glaser. 2000. Use of computer technology to help students with special needs. *The future of children* (2000), 102–122.
[16] Anthony J. Hornof and Linda Sato. 2004. EyeMusic: Making Music with the Eyes. In *NIME*.
[17] Thomas Theodore Scott Ingram and Jane Barn. 1961. A description and classification of common speech disorders associated with cerebral palsy. *Developmental Medicine & Child Neurology* 3, 1 (1961), 57–69.
[18] Sara K Jones. 2015. Teaching students with disabilities: A review of music education research as it relates to the Individuals with Disabilities Education Act. *Update: Applications of Research in Music Education* 34, 1 (2015), 13–23.
[19] Wojciech Kulak, Wojciech Sobaniec, Joanna Smigielska-Kuzia, Bozena Kubas, and Jerzy Walecki. 2005. A comparison of spastic diplegic and tetraplegic cerebral palsy. *Pediatric neurology* 32, 5 (2005), 311–317.
[20] Jeppe Veirum Larsen, Dan Overholt, and Thomas B. Moeslund. 2016. The Prospects of Musical Instruments For People with Physical Disabilities. In *Proceedings of the International Conference on New Interfaces for Musical Expression (2220-4806, Vol. 16)*. Queensland Conservatorium Griffith University, Brisbane, Australia, 327–331. http://www.nime.org/proceedings/2016/nime2016_paper0064.pdf
[21] Alex Michael Lucas, Miguel Ortiz, and Franziska Schroeder. 2019. Bespoke Design for Inclusive Music: The Challenges of Evaluation.. In *NIME*. 105–109.
[22] Richard E Mayer, Kristina Sobko, and Patricia D Mautone. 2003. Social cues in multimedia learning: Role of speaker's voice. *Journal of educational Psychology* 95, 2 (2003), 419.
[23] Joan McComas and Heidi Sveistrup. 2002. Virtual reality applications for prevention, disability awareness, and physical therapy rehabilitation in neurology: our recent work. *pre* 26, 2 (2002), 55–61.
[24] World Health Organization et al. 2011. *World report on disability 2011*. World Health Organization.
[25] Ricardo Pagán-Rodríguez. 2014. How do disabled individuals spend their leisure time? *Disability and health journal* 7, 2 (2014), 196–205.
[26] Kwang-Hyun Park, Zeungnam Bien, Ju-Jang Lee, Byung Kook Kim, Jong-Tae Lim, Jin-Oh Kim, Heyoung Lee, Dimitar H Stefanov, Dae-Jin Kim, Jin-Woo Jung, et al. 2007. Robotic smart house to assist people with movement disabilities. *Autonomous Robots* 22, 2 (2007), 183–198.
[27] Samuel Thompson Parke-Wolfe, Hugo Scurto, and Rebecca Fiebrink. 2019. Sound Control: Supporting Custom Musical Interface Design for Children with Disabilities. In *Proceedings of the International Conference on New Interfaces for Musical Expression*, Marcelo Queiroz and Anna Xambó Sedó (Eds.). UFRGS, Porto Alegre, Brazil, 192–197. http://www.nime.org/proceedings/2019/nime2019_paper038.pdf
[28] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174033
[29] Shimon Sapir, Lorraine Ramig, and Cynthia Fox. 2008. Speech and swallowing disorders in Parkinson disease. *Current opinion in otolaryngology & head and neck surgery* 16, 3 (2008), 205–210.
[30] Tommaso Schirinzi, Andrea Sancesario, Enrico Bertini, Enrico Castelli, and Gessica Vasco. 2020. Speech and language disorders in Friedreich ataxia: highlights on phenomenology, assessment, and therapy. *The Cerebellum* 19, 1 (2020), 126–130.
[31] Graham Welch, Evangelos Himonides, Jo Saunders, Ioulia Papageorgi, and Marc Sarazin. 2014. Singing and social inclusion. *Frontiers in psychology* 5 (07 2014), 803.