That's a Rap

Increasing Engagement with Rap Music Performance by Virtual Agents

Stefan Olafsson(✉), Everlyne Kimani, Reza Asadi, Timothy Bickmore

College of Computer and Information Science, Northeastern University, Boston, MA
{stefanolafs, kimani15, asadi, bickmore}@ccs.neu.edu

Abstract. Many applications of virtual agents, including those in healthcare and education, require engaging users in dozens or hundreds of interactions over long periods of time. In this effort, we are developing conversational agents to engage young adults in longitudinal lifestyle health behavior change interventions. Hip-hop and rap are one of the most popular genres of music among our stakeholders, and we are exploring rap as an engagement mechanism and communication channel in our agent-based interventions. We describe a method for integrating rap into a counseling dialog by a conversational agent, including the acoustic manipulation of synthetic speech and accompanying character dance animation. We demonstrate in a within-subjects study that the participants who like rap music preferred the rapping character significantly more than an equivalent agent that does not rap in its dialog, based on both self-report and behavioral measures. Participants also found the rapping agent significantly more engaging than the non-rapping one.

Keywords: relational agent, young adults, rap music, health counseling.

1 Introduction

Many applications of virtual agents, in areas such as health behavior change and education, require maintaining user engagement and retention over long periods of time and dozens or hundreds of interactions. Several approaches to maintaining engagement with virtual agents have been explored, including storytelling [1], automated camera motion [2], language variability, agent back-stories, and empathetic language use [3].

Music, including singing and dancing by a virtual agent, represents a relatively unexplored modality for user-agent interaction, but one that has the potential for significantly increasing user engagement. Most people find some form of music entertaining and this may add perceived value to interactions with an agent, thereby increasing the likelihood of user commitment to their relationship with the agent (according to the investment model of personal relationships [3]). "Entertainment education", using a variety of conventional media, has also been used as a successful health behavior

adfa, p. 1, 2011.

change modality [4], indicating that music may be an effective channel for health education.

Music has been used successfully in behavior change interventions [5] and could be used in a variety of ways in agent-based systems. For example, it can be a source of entertainment to further engage the user in conversations and for the agent to tell its backstory in an effort to build a relationship with the user. Through this medium, an agent can communicate health messages directly and succinctly. Lyrics provide the means to tell a story from multiple perspectives and for motivating behavior change, for example, through raising awareness of the importance of change and helping users reappraise the repercussions of inaction for other people and themselves.

We hypothesize that an agent that performs music in its interaction with users will increase user engagement and a higher likelihood of user retention in longitudinal interventions, at least for users who enjoy the particular genre of music being performed. In our current work, we are developing longitudinal health behavior change interventions for young adults. Since hip-hop and rap music are some of the most popular music genres among young adults (ages 18-25) in the world [6], there is reason to believe that this particular genre of music may be effective at boosting engagement for this demographic.

In this initial work we are exploring the use of rap music both to engage users and to establish backstory for the agent.

2 Related Work

Researchers have explored incorporating rap music into their (non-automated) health interventions, e.g., for substance use risk awareness [7]. In the fields of social work and psychotherapy, researchers have looked at engaging urban youth using rap music [8, 9]. DeCarlo and Hockman compared adolescents’ perception of the usefulness of integrating rap music as a tool to support prosocial skills and traditional group therapy. The results showed that a large percentage of the participants were in favor of the rap therapy compared with the traditional therapy [10]. Additionally, Olson-McBride and Page showed that hip-hop and rap music promoted self-disclosure and self-expression in high-risk youths [11].

Several systems have been developed to automate some aspect of music composition or performance by a virtual agent, e.g., automating facial expression [12], conducting an orchestra [13], dancing [14], and lyric generation [15]. However, there is little empirical evidence that music performance increases engagement with virtual agents. To our knowledge, this is the first study that evaluates the use of music in an agent based interactions in the context of health education and counseling.

3 Design of a Rapping Health Counselor Agent

To explore the feasibility and acceptance of having a virtual agent health counselor perform rap music, we developed two virtual agents that counseled users on two different health topics, with or without a rap performance.

We designed three dialogues for the agents. The first two were brief health counseling dialogues motivating exercise and good nutrition. These dialogues also included greetings, extended social chat, and farewells. During the extended social chat, the agent performed the rap. The third dialogue was an extended social chat intended to test engagement in the evaluation study.

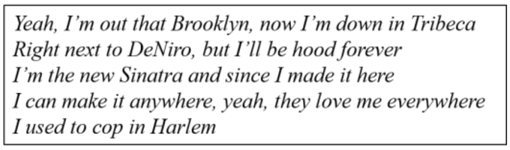
The agents in our study performed two short verses of rap from the popular 2009 rap and hip-hop song “Empire State of Mind” by the rapper Jay-Z, featuring Alicia Keys. This song was chosen because of its popularity [17] and used as part of the agents’ backstory. Figure 1 shows the transcription of the first verse.

The agents are animated in a 3D game engine (Unity 3D), speak using synthetic speech (Windows 10 TTS), and use a range of conversational nonverbal behavior for their non-singing utterances, including hand gestures, posture shifts, head nods, and eyebrow raises. Interaction with the agents is driven using a hierarchical transition network-based dialogue engine, with user inputs to the conversation selected from a multiple-choice menu updated at each turn of the conversation (Figure 2). Agent nonverbal behavior for non-singing utterances is automatically generated using BEAT [16] and synchronized with the synthetic speech. The animation accompanying the music is manually created and consisted of the agent bobbing its head, facial expressions, and lip-sync.

To generate the agents’ rap performance, the speech synthesizer was first used to record audio of the agent speaking the lyrics. We then manipulated the timing and duration of the words and the pitch contour of this speech sample.

To change the word durations, we needed to know the exact timings of the words in both original rap song and the synthesized sample. We used the “SPeech Phonetization Alignment and Syllabification” (SPPAS) toolkit [18] to align the song lyrics with voice samples. First, we performed Inter-Pausal Units (IPUs) segmentation to segment the audio signal into units of speech bounded with pauses of at least 300 milliseconds length. Each IPU was aligned with the corresponding segment of the lyrics. Then tokenization or “Text Normalization” was performed to remove punctuation, convert numbers and symbols to written forms, and segment text into words. Finally, words were converted into phonemes that were aligned with speech signal. SPPAS uses the Julius speech recognition engine [19] and HTK acoustic models trained from 16000 Hz audio samples to perform alignment. We aligned the song lyrics to both the original song and the synthesized voice.

We used Praat [20] to manipulate the synthesized sample word timings and pitch contour. The start and end times for each word in synthesized speech and original song were extracted from the SPPAS alignment outputs. The word timings in the synthesized sample were modified to match the corresponding times in the original song. We also extracted the pitch contour of the original song, which was used as the reference for manual modification of the pitch contour of the synthesized voice.



Yeah, I'm out that Brooklyn, now I'm down in Tribeca
Right next to DeNiro, but I'll be hood forever
I'm the new Sinatra and since I made it here
I can make it anywhere, yeah, they love me everywhere
I used to cop in Harlem

Fig. 1. Transcript of the agent's first rap verse from 'Empire State of Mind'.

4 Evaluation Study

We conducted an empirical study to determine the impact of a health counseling agent performing rap in its dialog. The study was a randomized, counterbalanced, within-subjects experiment with two treatments: R, where an agent performs rap as a part of the conversation; and NoR, where no rap was performed. In addition to randomizing the order of the experimental treatments, we used two different agents (Figure 2) that discussed two health topics, either nutrition (N) or exercise (E). Thus, our conditions were RN, RE, NoRE, and NoRN. These were randomized such that the assignment of topic to experimental treatment was counterbalanced across subjects.

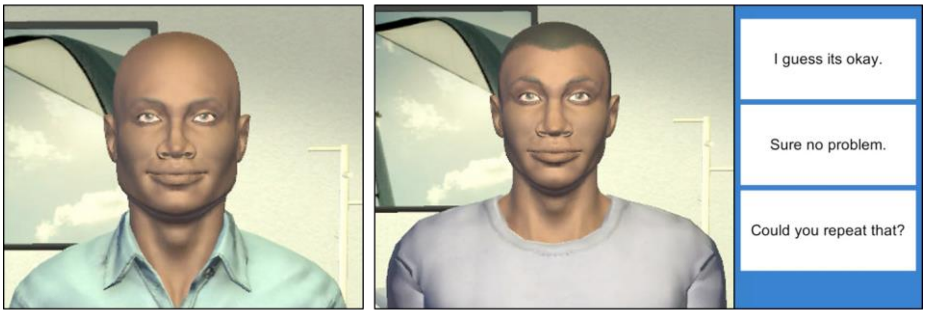


Fig. 2. The two conversational agents used in the study, with an example of user dialog options with the agent on the right.

4.1 Procedure

Participants were first asked to give informed consent, fill out a demographics questionnaire, and a questionnaire assessing their attitude towards rap music (see section 4.2). They were randomized using blocked randomization into either RE, RN, NoRE, or NoRN for their first conversation. The second conversation, therefore, consisted of the opposite configuration, e.g., if the first condition was RE, then the second would be NoRN. Following each of the first two conversations, participants were asked to fill out questionnaires regarding their experience of the agents (see section 4.2).

At this point participants were told that they would have a third conversation and that they could choose which agent they preferred to interact with. After selecting their preferred agent, the third conversation was launched. The content of this conversation was the same for all participants, consisting of 40 turns of social chat, with topics ranging from reading books to TV and movies, and users had the option to end the conversation at each turn after the 8th turn of dialog (with an option such as "Sorry, I have to go now").

Following the third conversation, we conducted a short semi-structured interview with participants about their experience with the agents.

4.2 Measures

At the start of the study, we collected socio-demographic information and assessed the degree to which participants liked rap music using the Rap Music and Attitude Perception (RAP) scale, a 25 5-Point scale item self-report questionnaire [21]. We also asked participants how often they listened to rap music and how much they liked the agent's rap, following their conversation with the rapping agent (Table 1).

We collected six measures of participant engagement. The first four were self-reported satisfaction, willingness to continue working with the agent, liking of the agent, and which agent participants felt was most engaging. The other two were behavioral measures: which agent the participant chose to have the third conversation with and the number of turns of talk the participant had with their chosen agent in the third conversation. Additionally, we collected a measure of trust in the agent.

The satisfaction, liking, and willingness to continue working with the agent were assessed with 7-Point scale items (Table 1). The trust questionnaire was a 15 item semantic-differential scale [22]. These assessments were obtained following each of the first two conversations.

The agent choice measure was collected after the second conversation by asking participants which agent they would prefer to interact with for a third and final conversation. The most engaging agent measure was collected by asking participants which agent they found most engaging, following all three conversations.

Participants reported on additional single 7-Point scale items touching on various aspects of their interaction with the agents (Table 1).

4.3 Participants

We recruited participants from a job listing website and by distributing flyers at various locations in Boston. Participants were required to be between 18-25 years old and be able to speak and read English. All participants were compensated $15 for their time. The study was approved by the University Institutional Review Board.

4.4 Evaluation Study Results

Participants. A total of 84 participants were enrolled, although one participant experienced a technical malfunction during the study session and had their data excluded. Participants were 60% male, aged 22.8 years (sd. 2.4), 79% Asian, 18% White, 2% Hispanic, and 1% Black, and 83% were students.

Table 1. Self-report single scale items (Wilcoxon Signed-Ranks tests, N=73).

Item (Anchor 1-7)	**When assessed**	**R avg (sd)**	**NoR avg (sd)**	**p**
How much did you like the agent's rap? (Not at all – Very much)	After Rap agent conversation	5.41 (1.6)	N/A	N/A
How satisfied are you with the agent? (Not at all – Very satisfied)	After each conversation	5.68 (1.2)	5.49 (1.3)	0.12
How much would you like to continue working with the agent? (Not at all – Very much)	After each conversation	5.52 (1.4)	5.32 (1.6)	0.19
How much do you like the agent? (Not at all – Very much)	After each conversation	5.63 (1.2)	5.47 (1.3)	0.16
How knowledgeable was the agent? (Not at all – Very knowledgeable)	After each conversation	5.71 (1.1)	5.95 (1.1)	0.08
How natural was your conversation with the agent? (Not at all – Very natural)	After each conversation	5.14 (1.6)	5.15 (1.7)	0.98
How would you characterize your relationship with the agent? (Complete stranger – Close friend)	After each conversation	4.14 (1.5)	3.7 (1.6)	0.001
How similar do you feel that you are to the agent? (Very different – Very similar)	After each conversation	4.36 (1.6)	4.26 (1.5)	0.43

Engagement. When all participants were included in the analysis, there were no significant differences between experimental treatments on engagement measures. We found that 13% of participants indicated they did not like rap music (scoring below 3 on the RAP scale). Thus, in the following analyses these participants are excluded.

Participants who like rap choose the agent in the R conditions significantly more often for the third conversation, (62% vs. 38%), $X^2(1)=3.96$, p<.05. When they were asked during the exit interview which agent was more engaging, they mentioned the R agent significantly more often than the NoR agent, (65% vs. 35%) $X^2(1)=6.21$, p<.05.

There were no significant differences in self-reported satisfaction between the two conditions (5.68 for R, 5.49 for NoR), Wilcoxon Signed-Ranks test W=284, n.s. However, trust in the NoR agent was rated significantly higher compared to the R agent, (6.64 for Rap, 6.81 for NoR), W=1501, p<.05. Desire to continue working with the agent was not significant between the two conditions, W=348.5, n.s.

Participants reported having a significantly closer relationship with the R agent compared to the NoR agent (scoring 4.1 vs. 3.7), W=184, p<.001. No significant differences were found between the groups on naturalness (W=565, n.s.), knowledgeability (W=556, n.s.), perceived similarity (W=606, n.s.), or liking of the agent (W=345, n.s.).

The number of dialogue turns in the third social conversation was not significantly different depending on the agent chosen, 22 for R, 25 for NoR, W=658.5, n.s.

Correlations. We conducted an exploratory bivariate correlational analysis of all quantitative measures, in order to understand how these factors influence one another. All were tested using Spearman's non-parametric rank order correlation.

There is a strong correlation between declaring that the R agent is most engaging and choosing the R agent for the third conversation (rho=.82, p<.001). There is a weak positive correlation between choosing the R agent and satisfaction with the R agent (rho=.22, p<.001), but a weak negative correlation between choosing the R agent and satisfaction with the NoR agent (rho=-.25, p<.05). Liking the agent's rap is weakly positively correlated with choosing the R agent (rho=.27, p<.05). Knowledgeability is also positively correlated with agent trust ratings (rho=.48, p<.001).

Desire to continue working with the agent was strongly correlated with liking the agent (rho=.81, p<.001), agent satisfaction (rho=.82, p<.001), naturalness of the conversation (rho=.63, p<.001), and perceived relationship with the agent (rho=.6, p<.001).

Perceived similarity to an agent is positively correlated with the perceived relationship with the agent (rho=.64, p<.001), agent satisfaction (rho=.41, p<.001), and trust, (rho=.45, p<.001). Perceived similarity to the R agent is also weakly, positively correlated with liking the agent's rap (rho=.31, p<.001).

4.5 Qualitative Interview Analysis

We interviewed participants at the end of the session and asked for their thoughts on the experience of interacting with the agents. Specifically, they were asked about their impressions of each interaction, which agent they found most engaging, and why. The interviews were transcribed and analyzed to identify concepts, which were then used to form themes. Two main themes specific to the agent interactions emerged: *Relatability & Engagement* (RE) and *Information & Knowledge* (IK). Concepts within the RE theme came up more frequently when participants were describing the R agent. Conversely, descriptions related to the IK theme were more frequent in the context of the NoR agent (Table 2).

Table 2. Concepts extracted from the semi-structured interviews by condition, frequency of occurrance, and theme. RE="Relatability and Engagement", IK="Information and Knowledge". Concepts with a frequency < 5 are not included in the table.

Rap (R)	**Frequency**	**Theme**	**No Rap (NoR)**	**Frequency**	**Theme**
Relatable	14	RE	Informative	20	IK
Engaging	12	RE	Engaging	7	RE
Interesting	10	RE	Knowledgeable	5	IK
Informative	8	IK	Friendly	5	RE

There was a distinction in how participants spoke about the agents. A greater number spoke about the NoR agent as one they learned something from, while more used adjectives like ‘fun’ or ‘entertaining’ when talking about the R agent. Some participants spoke of engagement and relatability together, as if finding the agent engaging depended on how well they could relate to it.

Finding the agent relatable and interesting came up particularly often in the context of the R agent, specifically when describing various aspects of the interaction, such as the conversation flow. Participants often described the agent in the NoR condition as informative and knowledgeable. Their prior knowledge and interest in health topics seemed to factor into their descriptions. For some participants, even if the agents didn’t provide them with any new information, they felt like the NoR condition had been more informative.

- *[I learned something from] the second one* [NoR condition] *because he talked a lot about the health and nutrition and the other one was a lot more fun* [R condition], *he had the music.* [#61 F 25]
- *The most engaging was the first one. It was more about, I could relate to him more that actually why I chose him for the third conversation and went with him.* [#17 M 19, R condition]
- *I could relate to it more like he was talking about music, rap music.* [#12 F 19, R condition]
- *It was more interesting flowing and it was like a friendly conversation.* [#10 M 23, R condition]
- *I learned some things so I would say the second was more informative.* [#13 F 19, NoR condition]
- *I didn’t learn anything new from either about the nutrition or the health but I feel like maybe from the first one.* [#37 M 23, NoR condition]

5 Conclusion

In this pilot study we evaluated a virtual agent that performed rap music in conversations with users about nutrition and exercise. Our results indicate that users who enjoy this particular genre of music find the agent that raps more engaging than one that does not and they are more likely to want to interact with that agent again. This demonstrates that rap music could be used to boost retention and adherence in longitudinal treatments with young adults who enjoy this music genre.

Young adults in our study who like rap music are likely to want to have more than one interaction with an agent that performs rap, especially if they feel like they can relate to the agent. The conversation that included the musical performance was designed to give a backstory for the agent and bring about a sense of relatability. Our qualitative and quantitative results show that one’s perceived similarity with the agent (relatability) was a particularly important determiner of ratings of trust in and satisfaction with the agent. This is in accordance with results from another study that found perceived similarity to be an important factor with respect to satisfaction with the agent [23].

Our results indicate that an agent that raps may have issues of credibility, given that the non-rapping agent was rated more trustworthy, described as more informative, and that trust and knowledgeability are significantly correlated. Desire to continue working with a particular agent was related to the perceived naturalness of the conversation, perceived relational closeness, liking, and satisfaction with the agent.

6 Future Work

This preliminary study is only a first step in our exploration of music in virtual agent interactions. There are numerous avenues for future investigations including, e.g., longitudinal applications. A variety of questions could also be studied as to the utility of embedding music within these applications, e.g., whether or not delivering health advice within rap lyrics increases knowledge retention.

Fully automated generation of singing and dancing behavior by a virtual agent is still an open problem, as is fully automated generation of rap lyrics to satisfy particular communicative and relational goals (beyond just coherence and entertainment goals, as targeted in [15]).

Ultimately, agents that dance and perform music must be evaluated in the context of longitudinal studies to determine, not only knowledge retention of content delivered through song, but the overall efficacy of health behavior change interventions incorporating these agents.

Acknowledgements. We thank Elise Masson, William Bond, Arsalan ul Haq, and Asimina Ino Nikolopoulou for their help designing the dialogue and conducting the pilot study.

References

1. Battaglino, C., Bickmore, T.: Increasing the Engagement of Conversational Agents through Co-Constructed Storytelling. In: Eighth Workshop on Intelligent Narrative Technologies (2015).
2. Ring, L., Utami, D., Olafsson, S., Bickmore, T.: Increasing Engagement with Virtual Agents Using Automatic Camera Motion. In: International Conference on Intelligent Virtual Agents (2016).
3. Bickmore, T., Schulman, D., Yin, L.: Maintaining Engagement in Long-Term Interventions With Relational Agents. Appl. Artif. Intell. 24, 648–666 (2010).
4. Hoffman, A.S., Lowenstein, L.M., Kamath, G.R., Housten, A.J., Leal, V.B., Linder, S.K., Jibaja-Weiss, M.L., Raju, G.S., Volk, R.J.: An entertainment-education colorectal cancer screening decision aid for African American patients: A randomized controlled trial. Cancer. 123, 1401–1408 (2017).
5. Lemieux, A.F., Fisher, J.D., Pratto, F.: A music-based HIV prevention intervention for urban adolescents. Heal. Psychol. 27, 349–357 (2008).
6. Hip-hop is the most listened to genre in the world | The Independent,

http://www.independent.co.uk/arts-entertainment/music/news/hip-hop-is-the-most-listened-to-genre-in-the-world-according-to-spotify-analysis-of-20-billion-10388091.html. Date accessed: 4/23/17

7. Paukste, E., Harris, N.: Using rap music to promote adolescent health: pilot study of VoxBox. Heal. Promot. J. Aust. 26, 24 (2015).
8. Levy, I.: Hip hop and spoken word therapy with urban youth. J. Poet. Ther. 25, 219–224 (2012).
9. Elligan, D.: Rap therapy: a culturally sensitive approach to psychotherapy with young African American men. J. African Am. Men. 5, 27–37 (2000).
10. DeCarlo, A., Hockman, E.: RAP Therapy: A Group Work Intervention Method for Urban Adolescents. Soc. Work Groups. 26, 45–59 (2004).
11. Olson-McBride, L., Page, T.F.: Song to Self: Promoting a Therapeutic Dialogue with High-Risk Youths Through Poetry and Popular Music. Soc. Work Groups. 35, 124–137 (2012).
12. Mancini, M., Bresin, R., Pelachaud, C.: A virtual-agent head driven by musical performance. In: IEEE Transactions on Audio, Speech, and Language Processiong. pp. 1883–1841 (2007).
13. Reidsma, D., Nijholt, A., Bos, P.: Temporal interaction between an artificial orchestra conductor and human musicians. Comput. Entertain. 6, 1 (2008).
14. Reidsma, D., Nijholt, A., Poppe, R., Rienks, R., Hondorp, H.: Virtual rap dancer. In: CHI ’06 extended abstracts on Human factors in computing systems - CHI ’06. p. 263 (2006).
15. Malmi, E., Takala, P., Toivonen, H., Raiko, T., Gionis, A.: DopeLearning: A Computational Approach to Rap Lyrics Generation. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’16. pp. 195–204 (2016).
16. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: BEAT: the Behavior Expression Animation Toolkit. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques - SIGGRAPH ’01. pp. 477–486. ACM Press, New York, New York, USA (2001).
17. Jay-Z, Keys’ “Empire” Tops Hot 100 For Fifth Week | Billboard, http://www.billboard.com/articles/news/266366/jay-z-keys-empire-tops-hot-100-for-fifth-week. Date accessed: 4/23/17
18. Bigi, B.: SPPAS: a tool for the phonetic segmentations of Speech. In: The eighth international conference on Language Resources and Evaluation. pp. 1748–1755 (2012).
19. Lee, A., Kawahara, T.: Recent Development of Open-Source Speech Recognition Engine Julius. In: APSIPA ASC 2009: Asia-Pacific Signal and Information Processing Association, 2009 Annual Summit and Conference. pp. 131–137 (2009).
20. Boersma, P.: Praat, a system for doing phonetics by computer. Glot Int. 5, 341–345 (2001).
21. Tyson, E.H.: Rap-music Attitude and Perception Scale: A Validation Study. Res. Soc. Work Pract. 16, 211–223 (2006).
22. Wheeless, L.R., Grotz, J.: The Measurement of Trust and Its Relationship to Self-Disclosure. Hum. Commun. Res. 3, 250–257 (1977).
23. Zhou, S., Bickmore, T., Paasche-Orlow, M., Jack, B.: Agent-User Concordance and Satisfaction with a Virtual Hospital Discharge Nurse. In: International Conference on Intelligent Virtual Agents (2014).