

Sonify - A Platform for the Sonification of Text Messages

Florian Alt¹, Bastian Pfleging¹

Institute for Visualization and Interactive Systems, University of Stuttgart¹

Abstract

Sonification of text messages offers a great potential for personalization while at the same time allowing rich information to be mediated. For example, ringtones are the major form of personalization on smartphones besides apps and background images. Ringtones are often used as a form of self-expression by the smartphone owner (e.g., using ones favorite sound track as standard ringtone), but also to identify the caller or sender of a message (e.g., the user knows who is calling without taking the phone out of the pocket). We believe this approach to be applicable to a wide variety of text messages, such as SMS, email, or IM. In this paper, we first present a web-based platform that allows user-generated mappings for text sonification to be created and managed. An API enables any application to send a text message and receive the sonification in the form of a MIDI file. To showcase the potential, we implemented an Android app that sonifies incoming SMS. Second, we evaluate the feasibility of our approach and show that sonified messages are equally effective as ringtones when conveying meta information.

1 Introduction

On polyphonic mobile phones, ringtones have long since become a central feature for personalization. Primarily, ringtones are used as a way of self-expression as well as to distinguish between different communication types (e.g., calls, SMS, or email). Incoming messages and calls can be easily linked to (groups of) contacts by assigning different ringtones. Currently, these ringtones are based on fixed audio files, such as (corporate) sounds and songs or personal recordings. The expressiveness of such ringtones is limited and the playback cannot adapt to the context of a message. However, it has been shown that audio can convey richer information, including the exact content of a message (e.g., Morse code). In addition, more abstracted information can be communicated, e.g., whether the message contains a question.

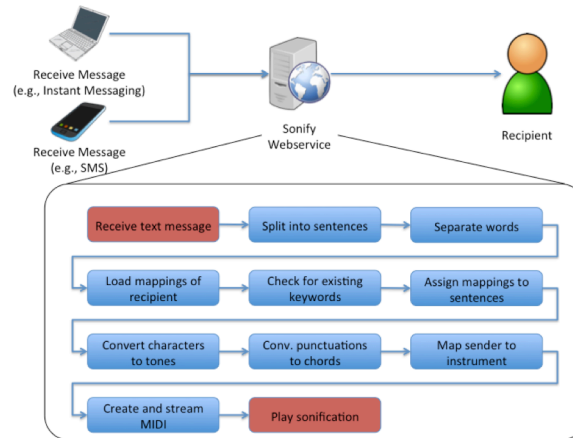


Figure 1: The sonification steps of our platform: an incoming message is first analyzed for keywords and then split into single characters. These characters are then mapped to tones, based on the keywords. Using different instruments, the sender can be encoded. Finally, a MIDI file is created from the sonification.

In this paper we present a platform that (1) extends this approach to arbitrary client applications dealing with sending and receiving messages and (2) enables users to generate personalized sonifications. We implemented the platform as a web-based service. External applications can send text messages using HTTP requests and obtain the sonification as raw data or MIDI file. Furthermore, it provides an interface for generating text-to-tone mappings. For encoding message information, the platform currently uses – but is not limited to – tonality (e.g., major, minor) and instruments (e.g., piano, violin). Further properties could include notes, rhythm, or velocity. In this way, not only expressive tones can be created, but users can also benefit from not having to immediately check a message’s content anymore.

Interesting use cases for our approach include (1) mobile phones, which users carry with them throughout the day, (2) messaging clients used in situations where users are often engaged in other activities (e.g., at work/during sports) to make them aware of important information in a non-obtrusive, yet understandable way, (3) assistance for users with impaired view, and (4) cases where users want to personalize their phones as a form of self-expression while at the same time benefitting from additional information. The contribution of this paper is twofold:

- We present a platform that allows different mappings to be created using a web interface. These mapping can be created either by users or by (professional) composers. The platform provides an API that makes these mappings available for arbitrary clients, converts text messages into a musical representation, and sends the representation back to the requester.
- Based on an online study we show that it is possible to encode multiple types of (meta) information into the sonification. In particular we tested the approach with the origin of a message as well as with the intention.

2 Related Work

Communication of information in non-verbal ways has been researched for over 150 years. One of the most prominent examples was the invention of the Morse code in the early 1840s. The popularity of this rhythm-based character encoding system stems from the fact that it can be read by humans without any decoding device and due to its high learnability.

More recently, sonification – the use of non-speech audio to convey information – has been explored for different types of data. MUSART is a sonification toolkit, which produces musical sound maps to be played in real-time (Joseph & Lodha 2002). Walker et al. (2004) presented the Audio Abacus, an application for transforming numbers into tones. The Sonification Sandbox (Walker & Cothran 2003) allows auditory graphs to be created from data sets. An important application area is sonification for visually impaired users. In this context Petrucci et al. (2000) showed how to use sonifications in auditory web browsers to allow spatial information to be explored by means of an audio-haptic interface.

Furthermore, the use of earcons in computer UIs was explored by Hankinson (Hankinson & Edwards 1999). Brewster (Brewster et al. 1993) evaluated whether earcons provide effective means for communicating information.

Different research projects have focused on the sonification of synchronous and asynchronous messaging. Avrahami (Avrahami & Hudson 2004) augment a commercial IM client with a tool that estimates the type of instant messages and modifies the salience of those deserving immediate attention. The results reveal that modifying notifications can create a benefit for the users. Isaacs et al. (2002) introduced Hubbub, a sound-enhanced mobile instant messaging client, aiming at increasing background awareness by providing audio clues. Sahami et al. (2010) investigated the impact of abstracted audio preview of SMS. The authors report that this type of preview affects the reading and writing behavior of users – however the approach focuses on simple notification tones only. Shake2Talk (Brown et al. 2009) is a mobile messaging system that allows for creating messages through gestures and send them to other users via their mobile phones.

Finally, research looked into the aesthetics of sonification. Inspired by the AIM (Arts in Multimedia) project, Babble online (Hansen & Rubin 2001) sonifies browsing activity, trying to communicate information both clearly and in a well-composed and appealing way. Song (Song & Beilharz 2008) presented mapping strategies derived from an analysis of various sound attributes, allowing for better representation and access of information from complex data sets.

Prior work has shown different ways to sonify data for various applications. In contrast, we shift the power of creating sonifications to the user. We extended our initial sonification prototype (Pfleging et al. 2012) and show, that even with little musical knowledge, data can be sonified in a way that can convey information to the user in an understandable way. In order to do so we provide a platform that supports the user in creating mappings and enables arbitrary applications to sonify text-based data.

3 Key Idea and Concept

Prior work strongly focused on sonifications that are based on direct mappings between data and sound. In this way a direct translation of a message into tones can be achieved. Though this enables information to be precisely transmitted, the decoding process is usually difficult for humans and requires a lot of practice.

The novelty of our approach lies in creating a sonification that can be instantly made sense of as it conveys information beyond the ‘raw’ data. Such meta information may include the sender or the intention of a message. Much of this information can be extracted from the original message based on keywords. Sahami et al. (2010) showed that text messages, such as SMS, include a lot of recurring keywords, making our approach presented in this paper particularly useful for text messages such as SMS, instant messages, or emails. In addition we believe that allowing the sonification to be customized can further enhance the ease of use of our approach. In this way users could associate arbitrary sounds (e.g., a piano) with the meta information (e.g., their partner).

One challenge can be seen in the aesthetics of the sonification, which should be designed in a way that makes sonification sound pleasing to the user. Prior work investigated ways to create euphonic sonifications (Alt et al. 2010), yet with limited means for customization. Hence, our approach allows arbitrary mappings to be created to provide as much artistic freedom as possible.

Our sonification algorithm is depicted in Figure 1. Each user can create and use different mappings that allow different information to be distinguished: (1) Once a message is received, the text is split into sentences and words. (2) For each sentence, the system scans for keywords; the first keyword found determines the mapping to be used for this sentence. If no keyword is found, a user-defined default mapping is used. (3) For a specific mapping, each single character is mapped to a note according to the defined mapping. (4) Punctuation within a message is used to encode its intention (e.g., question) as chords. (5) Finally, using different instruments allows various senders to be distinguished.

4 Tools and Platform

The overall aim of our work is to provide a globally accessible sonification platform for arbitrary text messages and devices. In order to achieve this, we envision a highly flexible service that a user can be configure and customize. Implementing such a technology as a web service offers different advantages:

- The service can be integrated into different message clients (e.g., SMS/e-mail readers, instant messengers) and platforms with low effort. A message is forwarded to the service and the reply is a sonification.
- Sonification configurations (mappings) can be shared among different message sources (clients) for one user and even among users.

- Composers and artists get the opportunity to create and share (or sell) sound objects and sonification themes.
- Streaming the sonifications to the client application hinders the unauthorized distribution of paid content within the platform.

To make the functionality ubiquitously and easily available, we implemented a web-based platform to offer all mentioned services. The platform consists of a database, which stores both mappings and user preferences, the program logic for creating the sonification, and an API for external clients to access the platform. Additionally, the platform offers a composer for creating the mappings and a web interface to manage the user preferences.



Figure 2: Web interface to create mappings for sonifications.

4.1 Composer: Creating Mappings

In order to allow different mappings to be created, an AJAX-based visual composer can be used to map characters to different notes. Figure 2 shows the interface of the composer: Each column represents one character and can be assigned to one of the notes shown in the rows on the left which represent a C major diatonic scale. By holding down the mouse button, the assignment between character and tone can smoothly be created.

As creating mappings that sound pleasing to humans is challenging and requires musical knowledge, we support users by highlighting tones (blue), which form a pentatonic scale, if selected together. Such a pentatonic scale can simply be created by combining five quint-

related tones (e.g., C, D, E, G, and A). Their use is highly popular in western music as they create highly euphonic melodies. Besides the main task of mapping characters to notes, a title, information on velocity, and a default instrument need to be defined. Additionally, the author of a mapping can freely assign keywords to this mapping, which activate this mapping if they occur in a sentence. To test a mapping, the user / designer can enter a test phrase and apply the current mapping to this message.

4.2 API

The platform provides a very simple API for encoding messages. The requesting client sends an HTTP post request which passes six parameters: username, password, the message to be encoded, the preferred response type, the message type, and information about the sender of the message. The platform then returns either a MIDI file or a link to the file on the platform with almost no delay.

4.3 Web Client: Personalized Mappings

A web client serves two tasks: first, it can be used to *manage the mappings*. In a first step, the user can select mappings from a list for his account. In a later version, these mappings could, e.g., be bought for a small fee. In a second step, the user defines the mappings he wants to activate. These mappings are then used to sonify an incoming text message. One of the mappings serves as a *default mapping* in cases where no keywords are contained in a message (cf. section 'Key Idea and Concept'). Second, the user can *assign instruments* to the mappings. In this way, further meta-information, such as the message sender, can be encoded. Currently, the user can select from a list of pre-defined instruments. Further properties could be added in order to encode additional information in a later version (e.g., tonality, velocity, rhythm).

4.4 Mobile App

To showcase the potential of the platform we created an Android app (running as a background service) that sonifies incoming SMS. As it receives a new text message, a BroadcastReceiver extracts the incoming message and transmits the name of the sender as well as the text to our web service. The web service decomposes and encodes the message as a MIDI file and sends it back to the phone.

5 Evaluation

As a proof-of-concept we evaluated our approach in a user study with the aim to investigate whether people could extract meta-information, such as the identity of the sender or the message intention.

5.1 Setup

For the study we had a student (22, male, non-musician) create a pentatonic mapping using our composing tool. As meta-information we used (1) the sender, who was encoded in the used instrument (we used a piano, a xylophone, and strings) and (2) the intention of a message (we encoded the following messages: “*Are you available tonight?*” (question), “*Yes, see you later.*” (positive intention), and “*No, today does not work for me.*” (negative intention). The mappings were designed in a way, such that playing back a message on a major scale represented positive intentions, a mapping played back on a minor scale represented negative intentions, and a sept chord in the end represented a question. Standard ringtones were used as a baseline.

To reach as large an audience as possible, we designed the evaluation as a web-based study. We created a website where users were presented different sonified text messages and then had to decide which meta-information was encoded in the message (within subject design). Messages were prerecorded (Flash) to make sure the sonification sounded equal for every participant. We recruited the participants via mailing lists and Facebook.

5.2 Procedure

In an initial questionnaire, we first asked the users for demographics (gender, age, profession, musical knowledge).

5.2.1 Extracting Single Meta Information

In the first part, participants were presented 3 ringtones and 3 instruments that could be separately played back via a button as often as they wanted. They then had the task to exclusively assign one ringtone and one instrument each, respectively, to three people they frequently communicate with, e.g., their partner, their best friend, or a family member. After the mapping from persons to ringtones and instruments was finished, a training phase allowed the participants to practice guessing the right sender as often as they wished. We then presented a set of 8 audio files to them, each of which contained either a melody played with the instrument or a ringtone, in randomized order. Each set contained all 3 ringtones and melodies from all 3 instruments plus 2 random audio files in order to prevent users from counting which instruments or ringtones had already been presented. The duration of the melodies was around 6 seconds, for the ringtones around 4 seconds. For each melody or ringtone the users then had to decide which sender had sent the message.

5.2.2 Extracting Multiple Meta Information

In the second part participants were explained the different intentions (positive, negative, question) and how they are encoded (major pentatonic scale, minor pentatonic scale, sept chords). We provided examples as well as an arbitrary training phrase. Then, similar to the first part, we presented the participants a set of audio files, which they had to decode. The participants had to decide for each audio file (1) who the sender was and (2) what the intention was. Overall 9 different audio files were used (3 senders x 3 intentions). Similar to

the first study, we added 3 fake sonifications to prevent users from counting combinations. The order was again randomized and the melody duration was around 7 seconds. No ringtones were used in this part.

5.3 Results

In total, 55 people participated in the study (avg. age=25.9 years, 40 males, 16 musicians). First, we looked at how many participants can *correctly identify the sender*. We found that for our sonification 73.6 % of the participants give the correct answer (81.4 % for the ringtones; difference non-significant; chi square: (1, 250)=2.2, $p>0.05$). No differences between musicians and non-musicians and male and female were found.

As the participants had the opportunity to replay the melodies as often as possible, we looked at *how often the participant listen* (click the playback button), as this can be seen as a measure for the difficulty of the task. Overall, participants listen to the sonification 129 times (ringtones: 120 times). We found that 75.8% of the participants make their choice after they have once played back the sonification (89.1% for the ringtone). Furthermore, we looked at *how quickly users make the decision*. We found that for the sonifications, the average duration is 5.38 s, SD=6.50 s (ringtone: avg. duration=4.90 s, SD=3.26 s). Note that these numbers include the playback time of the melody and ringtone as we are not able to determine whether the users listened to the entire melody or not.

Overall we did not observe any significant differences between ringtones and sonification. Similarly, there were neither significant differences between each of the three ringtones nor between each of the three different sonifications with regard to correct answers, number of playbacks, and time required for making a decision.

For the second part of the study we again tested how many participants can *correctly identify the meta-information*. For the message sender, 73.0% of the participants give the correct answer, which is in line with the first part of the study. For the message intention 72.7% of the participants can correctly identify whether the intention was positive (71.9%), negative (70.0%) or a question (76.4%). However, there is quite a large increase in the duration until people made their decision (8.86s), which suggests an increase in cognitive load.

6 Limitations

Currently, popular songs are typical choices for ringtones. Songs of featured artists are used to identify callers but at the same time are statements of self-expression. With the use of mappings instead of pre-recorded songs the recognition effect of a song is lost. However, as composers offer characteristic mappings (e.g., a Hard Rock mapping), these may become a similar form of self-expression.

In its current form the approach is limited to single character-to-tone mappings, which requires a rather difficult learning process, as each mapping is different. Future versions could further abstract messages or provide easy-to-learn mappings.

We deliberately used simple keyword spotting to detect the intention of a sentence as a proof of concept. There is a body of work that looks into text mining and text understanding – however this was not at the center of our research but it should be considered for a real-world installation. Additionally, our current server-based generation of sonifications was implemented to facilitate sonifications for a multitude of systems. A future release might allow sonifications to be generated locally if mappings are transferred to the client. Thus, sonifications can be generated offline without sending messages to a server to preserve the users' privacy.

7 Conclusion

In this paper we presented a platform that allows (1) user-generated mapping for the sonification of text messages to be created and (2) arbitrary clients to sonify text messages using a web-based API. Both 'raw' information as well as abstract meta data can be conveyed. Sonification is not limited to a single mapping but users as well as composers can create and share their own mappings, hence catering to mobile phone users' needs for self-expression. To show the utility of the platform we implemented an Android app that runs in the background and provides sonifications for incoming SMS.

An initial web-based study did not reveal any significant differences in accuracy and the time required to extract meta-information between (state-of-the-art) ringtones and our sonification approach. This suggests that our approach could easily be used as a substitute for such standard ringtones allowing at the same time for conveying additional information. However, the increase in time it takes to make a decision about the information also suggests an increased cognitive load.

We believe the approach to be applicable to a wide variety of application domains (including SMS, email, instant messages, Twitter feeds, Facebook posts, etc.) and devices (smartphones, tablet, laptops, etc.) – in general for any text message that is of personal interest to a user. Due to the unobtrusive nature we believe the approach to also be applicable in public or semi-public settings. Future work could investigate the approach in the long term, focusing on extracting more fine-grained information, such as the literal message. Furthermore we are interested in the impact on user experience, which meta-information users find particularly valuable, and how this information is encoded best.

Acknowledgements

We thank Erhan Eroglu for his help with implementing the system and during setting up and conducting the user study.

This project was sponsored by the German Research Foundation (DFG) in the Cluster of Excellence in Simulation Technology (EXC 310/1) at the University of Stuttgart.

References

- Alt, F., Shirazi, A., Legien, S., Schmidt, A. & Mennenöh J. (2010). Creating meaningful melodies from text messages. In *Proc. of NIME 2010*. New York, NY, USA: ACM, p. 63–68.
- Avrahami D. & Hudson, S. (2004). Qna: augmenting an instant messaging client to balance user responsiveness and performance. In *Proc. of CSCW '04*. New York, NY, USA: ACM p. 515–518.
- Brewster, S. Wright, P. & Edwards A. (1993). An evaluation of earcons for use in auditory human-computer interfaces. In *Proc. of INTERACT '93*. New York, NY, USA: ACM, p. 222–227.
- Brown, L., Sellen, A., Krishna, R., & Harper R. (2009). Exploring the potential of audio-tactile messaging for remote interpersonal communication. In *Proc. of CHI '09*. New York, NY, USA: ACM, p. 1527–1530.
- Hankinson, J. & Edwards, A. (1999). Designing earcons with musical grammars. *ACM SIGCAPH Computers and the Physically Handicapped* (65, Sept. 1999): p. 16–20.
- Hansen, M. & Rubin, B. (2001). Babble online: applying statistics and design to sonify the Internet. In *Proc. of ICAD '01*, p. 10–15.
- Isaacs, E., Walendowski, A. & Ranganthan, D. (2002). Hubbub: A sound-enhanced mobile instant messenger that supports awareness and opportunistic interactions. In *Proc. of CHI '02*. New York, NY, USA: ACM, p. 179–186.
- Joseph, A. & Lodha, S. (2002). Musart: Musical audio transfer function real-time toolkit. In Nakatsu, R. & Kawahara, H. (Ed.): *Proc. of ICAD '02*. Kyoto, JP: ATR, p. 64–69.
- Petrucci, L., Harth, E., Roth, P., Assimacopoulos, A. & Pun, T. (2000). Websound: a generic web sonification tool, and its application to an auditory web browser for blind and visually impaired users. In Cook, P.R. (Ed.): *Proc. of ICAD '00*, p. 2–5.
- Pfleging, B., Alt, F. & Schmidt, A. (2012). Meaningful Melodies: Personal Sonification of Text Messages for Mobile Devices. In *Proc. of MobileHCI '12*. New York, NY, USA: ACM, p. 189–192.
- Sahami Shirazi, A., Sarjanoja, A., Alt, F., Schmidt, A. & Häkkinä, J. (2010). Understanding the impact of abstracted audio preview of SMS. In *Proc. of CHI '10*. New York, NY, USA: ACM, p. 1735–1738.
- Song, H. & Beilharz, K. (2008). Aesthetic and auditory enhancements for multi-stream information sonification. In *Proc. of DIMEA '08*. New York, NY, USA: ACM, p. 224–231.
- Walker, B., & Cothran, J. (2003). Sonification sandbox: A graphical toolkit for auditory graphs. In Brazil E., & Shinn-Cunningham B. (Ed.): *Proc. of ICAD '03*. p. 161–163.
- Walker, B., Lindsay, J. & Godfrey, J. (2004). The audio abacus: Representing a wide range of values with accuracy and precision. In Barrass S. & Vickers, P. (Ed.): *Proc. of ICAD '04*.

Contact Information

Florian Alt, Bastian Pfleging
Institute for Visualization and Interactive Systems, University of Stuttgart
Pfaffenwaldring 5a, 70569 Stuttgart, Germany
florian.alt@vis.uni-stuttgart.de, bastian.pfleging@vis.uni-stuttgart.de