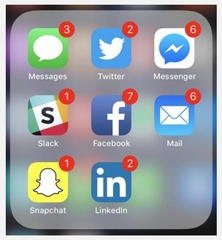
IEEE ICASSP 2021
Toronto, Canada
Paper ID: 2705

Unsupervised Musical Timbre Transfer for Notification Sounds

Jing Yang[1], Tristan Cinquin[1], Gábor Sörös[2]
[1]Department of Computer Science, ETH Zurich; [2]Nokia Bell Labs, Budapest, Hungary

1. The need for less intrusive notifications



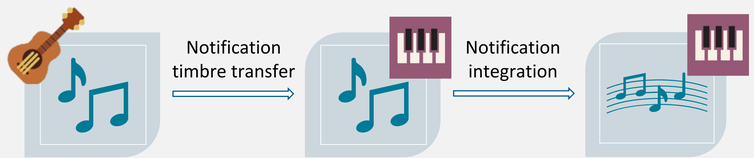
Notifications make people aware of surrounding activities [2] but can easily become a source of distraction and stress [1].

As digital music has been widespread and many people have the habit of conducting activities over music, there is potential to deliver notifications by altering the music the user is listening to.

How can we deliver the commonly used artificial notification sounds in a less intrusive way?

2. Calm notifications in adaptive styles

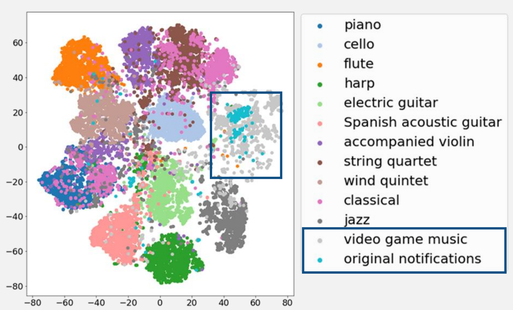
We propose a novel concept of delivering notification sounds by transforming the artificial notification sounds into musical timbres.



Step 1: Transfer the timbre of a notification sound into the timbre of music the user is listening to, while preserving the original notification melody.
Step 2: Deliver the timbre-transformed notification by harmoniously integrating it into the music.

3. Data collection

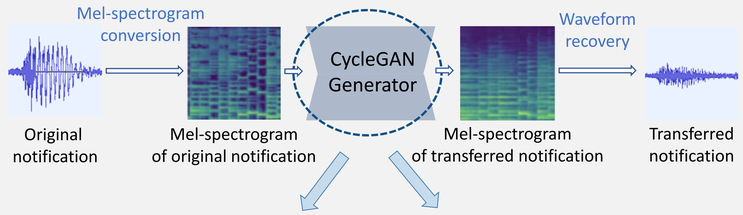
To tackle the issue of insufficient training data and ambiguous timbre of notification sounds, we collected video game music for source domain samples as enabled by proven timbral similarity.

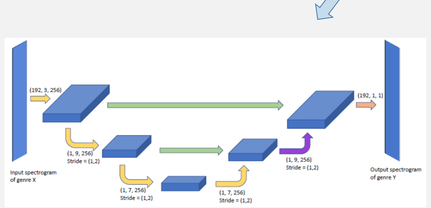


4. Musical timbre transfer for notifications

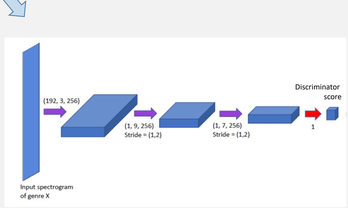
We change the musical timbre of artificial notification sounds

To tackle the lack of pairs of notifications and their corresponding music sequences, we perform unpaired audio style transfer using a CycleGAN-based model:



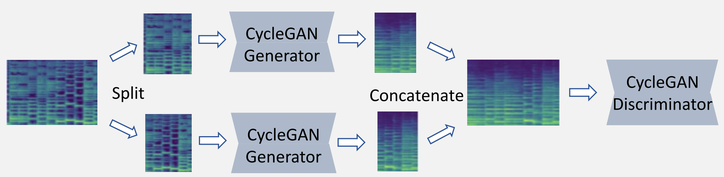


Generator:
fully convolutional U-Net architecture

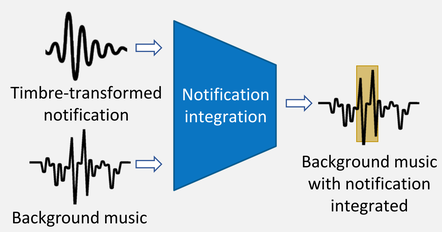


Discriminator:
convolutional PatchGAN discriminator

To process notification sounds of arbitrary length, we implement a splitting+concatenation mechanism [3]:

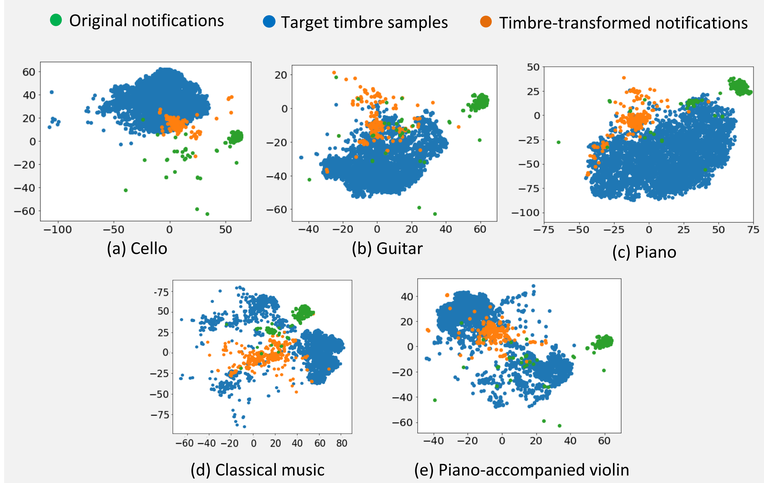


To embed notifications into music, we perform
1. Amplitude adjustment
2. Tempo adjustment
3. Fade-in and fade-out effects



5. Stylized notifications and delivery

Our method can well transfer single-timbre notifications into single-timbre styles (a,b,c), but cannot well transfer into multi-timbre styles (d,e).



Our method can preserve the original notification melody reasonably well. A user study with 53 participants gave a mean opinion score (MOS) of 3.720±0.261 on a 5-point scale for melody preservation.
Our method is offline but we would like to extend it to real-time adaptive styles.

Listen to **audio samples** at [4]!

We can replay the Nokia tune in arbitrary timbres

6. References

1. G. Mark, D. Gudith, U. Klocke – The cost of interrupted work: more speed and stress, ACM SIGCHI Conference on Human Factors in Computing Systems (CHI), 2008
2. S. T. Iqbal, E. Horvitz – Notifications and awareness: a field study of alert usage and preferences, ACM Conference on Computer Supported Cooperative Work (CSCW), 2010
3. M. Pasini – MelGAN-VC: Voice conversion and audio style transfer on arbitrarily long samples using spectrograms," arXiv:1910.03713, 2019
4. https://gladys0313.github.io/notification-timbre-transfer/