ETH zürich

DIStributed COmputing

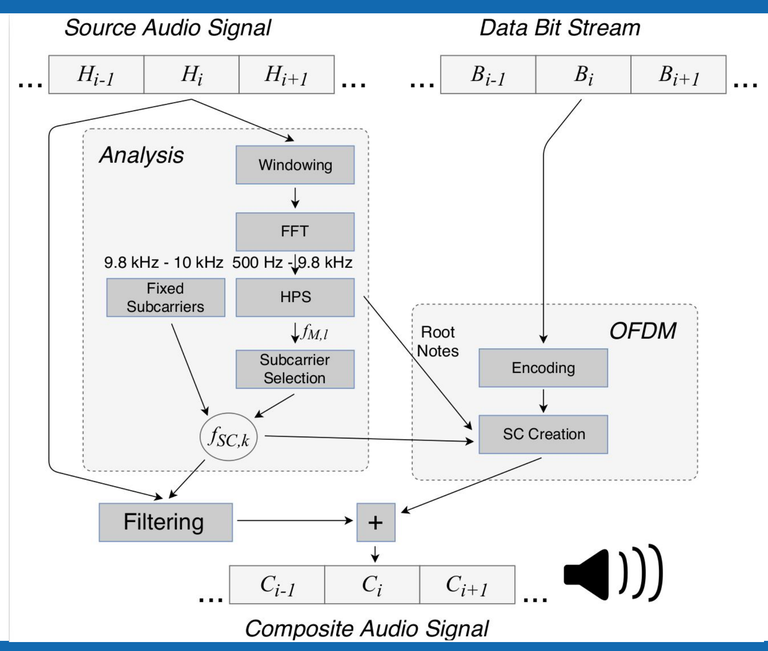
Imperceptible Audio Communication

Manuel Eichelberger, Simon Tanner, Gabriel Voirol, Roger Wattenhofer
Distributed Computing Group, ETH Zurich

1 Introduction

An off-the-shelf speaker plays an arbitrary piece of music with embedded data. A smartphone, recording with its microphone, can decode that data. We leverage the *masking effect*, a psycho-acoustic phenomenon which hides sounds that are close in time and frequency to a louder sound, a so-called *masker*. The music is temporally segmented and OFDM channels are placed in frequencies close to such maskers. Therefore, the added OFDM channels containing the data are imperceptible.

2 Overview

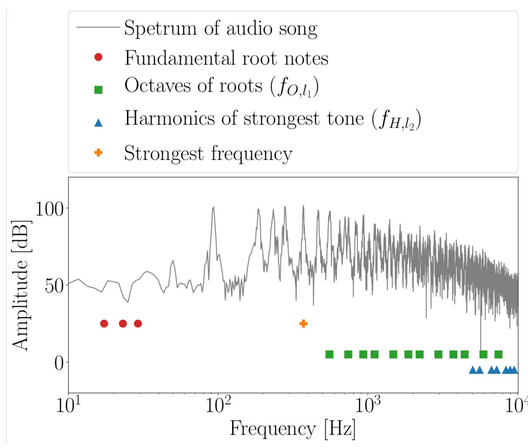


3 Method

- Audio frequency range shared between data and music
- Ultrasonic signals have low range and are directional. Therefore, we insert the data in lower, audible frequencies.
- Using the masking effect, our system transmits data in OFDM subcarriers next to frequencies of high amplitude.
- Due to the overtones of instruments, the harmonics (multiples) of loud notes also have high amplitude.

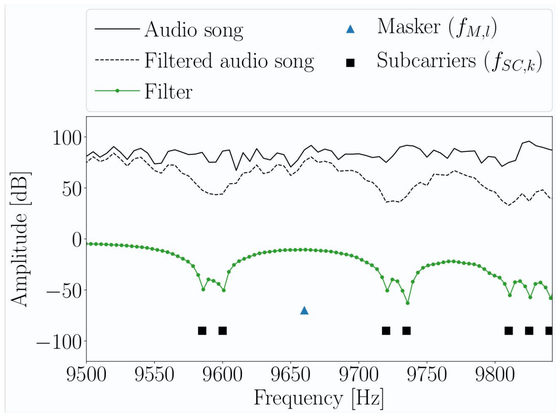
Analysis

- Compute Harmonic Product Spectrum (HPS) of song segments H_i
- Find fundamental root notes of the three most dominant tones in the harmonic chromatic scale between C_0 = 16.35 Hz and B_0 = 30.87 Hz
- Use all octaves of the three root notes and all harmonics of the strongest note up to 9.8 kHz as maskers



Data Insertion

- Four subcarriers close to each masker
- Fixed subcarriers from 9.8-10 kHz to transmit selection of root notes
- Dampen original song at subcarrier frequencies
- OFDM modulation of data bits

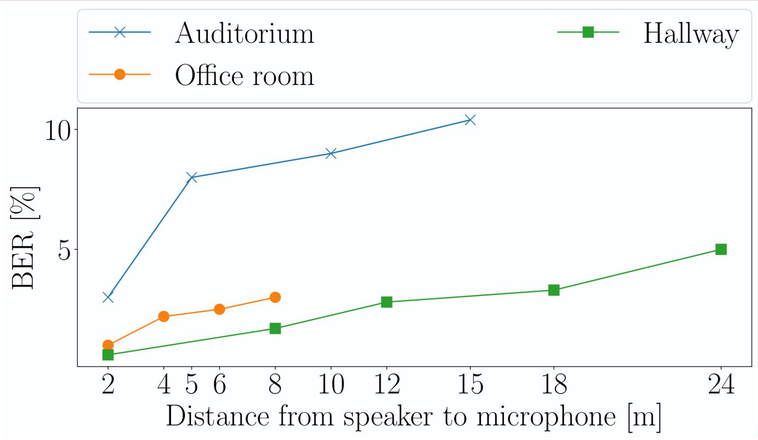


4 Results

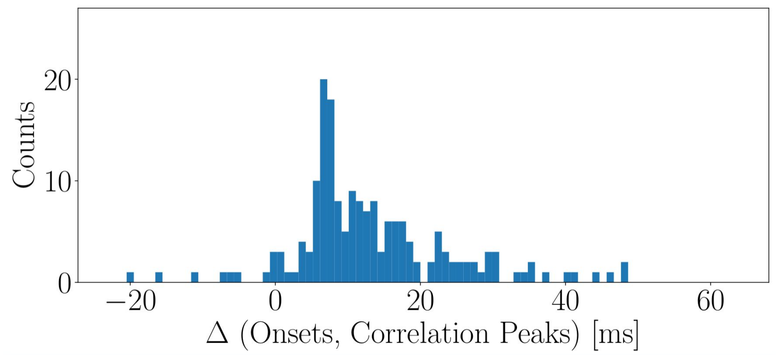
We play the modified music from a *KRK Rokit 8* speaker and record it with a *Nexus 5X* smartphone. The original music is encoded in the *Free Lossless Audio Codec (FLAC)*. Artifacts from lossy encoding might otherwise conceal noise introduced by the OFDM channels.

Our implementation embeds data at rates of 309-412 bit/s for different music styles. In a hallway, our system can transmit data up to 24 meters with bit error ratios (BERs) below 5%. In a big auditorium, the BER can be kept at 10% at a distance of 15 meters. The BERs in the table are measured at a distance of 2 m.

Artist – Song	Data Rate [bit/s]	BER [%]	Genre
Munstrous – When Am I (SourOne Remix)	400	4.2	Electronic
Van Halen – And The Cradle Will Rock	359	3.0	Classic Rock
Pink Floyd – Breathe	309	2.9	Ballade
Scorpions – Can't Live Without You	315	2.8	Classic Rock
Queen – The Show Must Go On	412	2.6	Ballade
Gorillaz – All Alone	324	2.2	Electronic



The histogram shows the time differences between the real and estimated OFDM symbol onsets in an auditorium at a distance of 5 m. Mostly, a valid onset inside the cyclic prefix from 0 to 66.6 ms is found. Our receiver uses the periodicity of the OFDM symbol onsets to make the detection more robust.



Subjective audio quality test

- Experiment 1 (E1): Participants hear either the original (O) or the modified (M) song and decide which one it is.
- Experiment 2 (E2): Participants hear both songs and decide which one is which. p(E) is the error (wrong assignment) probability.

Artist – Song	$p(O\|O)$, E1 [%]	$p(O\|M)$, E1 [%]	Δp, E1 [%]	$p(E)$, E2 [%]
Munstrous – When Am I (SourOne Remix)	40.7	71.4	-30.7	71.8
Van Halen – And The Cradle Will Rock	33.3	40.0	-6.7	38.5
Pink Floyd – Breathe	64.0	62.5	1.5	20.5
Scorpions – Can't Live Without You	92.0	87.5	4.5	23.1
Queen – The Show Must Go On	60.0	68.8	-8.8	53.8
Gorillaz – All Alone	33.3	40.9	-7.6	35.9
Average (40 Participants)	53.9	61.8	-7.9	40.6

5 Conclusion

Data can be hidden imperceptibly in background music, while allowing robust decoding. The computational effort for receiving the hidden data is low enough for real-time processing on smartphones.

Future Work

- Real-time receiver implementation in a smartphone app, maybe tailored for a specific use case.
- Additional user studies, to optimize the number and amplitude of the subcarriers to allow higher data rates.
- Data rates could be increased by adapting the music, for instance by amplifying maskers.