Aerial Communications using Piano, Clarinet, and Bells

Natacha Domingues*, João Lacerda*, Pedro M. Q. Aguiar† and Cristina V. Lopes‡

*Instituto Superior Técnico (IST), Lisboa, Portugal. E-mail: {natachadomingues,joaolacerda}@megamail.pt

†Institute for Systems and Robotics / IST, Lisboa, Portugal. E-mail: aguiar@isr.ist.utl.pt

‡ Xerox PARC, Computer Science Lab., Palo Alto CA, USA. E-mail: lopes@acm.org

***Abstract*— This work explores novel mechanisms for aerial acoustic machine-machine communications. It builds on previous work by some of the authors [1], as well as others [2]. In this paper we describe aerial acoustic communication systems that sound like musical instruments. The sound primitives come from simple models for the sound of the piano, the clarinet, and the bells. The messages are coded by combining these primitives according to musical harmony. Our experiments show that these communication systems are well suited for applications requiring very low bit rates. Examples of the acoustic signals produced are made available from the WWW.**

I. Introduction

Recently, some researchers, including ourselves, have revisited the use of sound for machine to machine communication, decades after the original acoustic modems were considered outdated [1], [2]. Sound has properties that makes it appealing for ubiquitous computing and personal area networks, namely the fact that it's a broadcast medium with very short range and localized propagation. Issues of privacy and security are addressed naturally. The audible audio band also has a unique advantage when the goal of transmitting data from one device to others is combined with the possibility of exposing some information to humans about the messages that are being transmitted. The fact that there exists a massive infrastructure for human voice enables no-cost solutions for linking traditional media, such as radio and TV, with computing devices.

One of the goals of our work is to explore the field of artificial communications when humans are also part of the communication system, from simple awareness to complete understanding of the message. A niche in this space is the use of musical sounds which are familiar and pleasant to humans.

Reference [1] shows how to obtain relatively pleasant sounds using standard modulation schemes such as *Frequency Shift Keying* (FSK) and *Amplitude Shift Keying* (ASK). Standard FSK assigns each symbol in the alphabet to a pure tone frequency. In this paper we combine standard FSK with musical sound synthesis in order to assign each symbol to a more pleasant musical sound. These musical sounds are chosen to resemble the sounds of the piano, the clarinet, and the bells. The resulting communicating systems sound like those musical instruments as messages are transmitted.

By mapping the symbols into musical notes in a careful way, we were able to decode the messages using a standard FSK demodulator. This way, although the spectra of the transmitted signals is complex, our receiver is kept computationally very simple. Our experiments demonstrate that this kind of aerial musical communications system is well suited for applications that require very low bit rates. Files with examples of the sounds produced are made available from the WWW link [3].

Paper organization

In section II we address the synthesis of musical sounds. Sections III and IV describe the modulation technique and the detection method used. Experimental results are reported in section V. Section VI concludes the paper.

II. Musical Sound Synthesis

The musical instruments we synthesized were the piano, the bells, and the clarinet. Our challenge was reproducing the main characteristics of the sound of those instruments in such a way that the synthesized acoustic waveforms could be used in a digital communication system without requiring specific and complex coding schemes and, consequently, very expensive receivers. To achieve this goal, we synthesized the audio spectra of the musical instruments by either explicitly choosing the spectral components of the sound (in the case of the piano) or by means of frequency modulation techniques [4], [5] (in the case of the bells and clarinet). To replicate the time-varying acoustic power of the signal, the synthesized waveforms are then windowed with an appropriate time envelope.

Vibrating strings – piano

A very simple model for the sound of the piano ignores the soundboard and simply replicates the sound of a vibrating string. To synthesize the vibrating string sound we used a periodic acoustic signal with fundamental frequency f_0. The piano sound is then given by

$$x(t) = A(t) \sum_{n=1}^{N} a_n \cos(2\pi n f_0 t), \qquad (1)$$

where the time envelope $A(t)$ is the *attack-delay-sustain-release* (ADSR) profile widely used to model string instruments [5]. The shape of the periodic waveform, or, equivalently,

0-7803-7713-3/02/$17.00 © 2002 IEEE

the amplitudes $\{a_n, 1 \le n \le N\}$ of its spectral components, determines the timbre of the sound. We used $N = 5$ spectral components with decreasing amplitudes, chosen by experimentation, $a_1 = 1$, $a_2 = 0.5$, $a_3 = 0.3$, $a_4 = 0.2$, $a_5 = 0.1$.

Vibrating rigid metals and air columns

To synthesize the more complex spectra of both the sound of vibrating rigid metals, such as the plates or the bells, and the sound of vibrating air columns, we use a model based on frequency modulation techniques as proposed in [4], see also [5]. According to this model, the acoustic signal is a modulated sinusoid

$$x(t) = A(t)\cos\big(\phi(t)\big), \tag{2}$$

where $A(t)$ is the time envelope and the phase $\phi(t)$ is given by

$$\phi(t) = 2\pi f_c t + I(t)\cos\left(2\pi f_m t + \Phi_m\right) + \Phi_c, \tag{3}$$

where f_c is the carrier frequency, f_m is the modulating frequency, $I(t)$ is the modulation index envelope, and Φ_m and Φ_c are phase constants that we set to $\pi/2$ to avoid signal discontinuities.

To provide an intuitive interpretation for the phase $\phi(t)$ in (3), we compute the instantaneous frequency of the acoustic signal $x(t)$ in (2). By differentiating the phase (3), after simple manipulations, we express the instantaneous frequency $f_i(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt}$, as

$$f_i(t) = f_c - \Re\left\{\left(I(t)f_m - j\frac{dI(t)}{dt}\frac{1}{2\pi}\right)e^{j2\pi f_m t}\right\}, \tag{4}$$

where $\Re\{z\}$ stands for the real part of the complex z. Expression (4) shows that the instantaneous frequency $f_i(t)$ oscillates around the constant frequency f_c. For example, if the modulation index $I(t)$ is constant, the instantaneous frequency $f_i(t)$ is a sinusoid with mean value f_c and frequency f_m. Depending on $A(t)$, $I(t)$ and the ratio f_c/f_m, we obtain an acoustic signal $x(t)$ that resembles the sound of a bell or a clarinet.

Bells

Choosing both the time and the modulation index envelopes as decaying exponentials, $A(t) = A_0\exp(-\frac{t}{\sigma})$, $I(t) = I_0\exp(-\frac{t}{\tau})$, and the ratio $f_c/f_m = 0.5$, we synthesized the bells sounds [4], [5]. To illustrate the physical meaning of these parameters, we plot in Figure 1 the signal $x(t)$ and its instantaneous frequency $f_i(t)$, for $A_0 = 1$, $I_0 = 10$, $f_c = 5$Hz, $f_m = 10$Hz, and $\sigma = \tau = 0.5$ (the frequency values were chosen for illustration purposes only).

The left plot of Figure 1 shows that the time envelope of the acoustic signal decays exponentially as expected. From the right plot of Figure 1, we see that the envelope of the instantaneous frequency is also a decaying exponential. In fact, by particularizing the instantaneous frequency (4) to this case, we get

$$f_i(t) = f_c - I_0 e^{-\frac{t}{\tau}}\,\Re\left\{\left(f_m + j\frac{1}{2\pi\tau}\right)e^{j2\pi f_m t}\right\}, \tag{5}$$

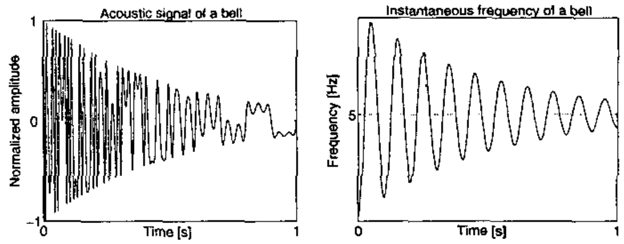

Fig. 1. Bell. Left: time evolution of the acoustic signal. Right: its instantaneous frequency.

which shows that the instantaneous frequency $f_i(t)$ of the bell sound has an underdamped oscillation – the amplitude of the oscillation decays exponentially, its mean value is f_c, and its damped frequency of oscillation is f_m. This is in agreement with the right plot in Figure 1.

Clarinet

Defining the envelopes as the composite functions

$$A(t) = \begin{cases} e^{\frac{t}{a}\ln 2} - 1 & t \in [0, a] \\ 1 & t \in a + [0, s] \\ 2 - e^{\frac{t-a-s}{r}\alpha} & t \in a+s+[0, \frac{r}{2}] \\ e^{-\frac{t-a-s-r}{r}\alpha} - 1 & t \in a+s+\frac{r}{2}+[0, \frac{r}{2}], \end{cases} \tag{6}$$

where the constant $\alpha = 2\ln 1.5$ makes $A(t)$ continuous, and

$$I(t) = \begin{cases} 5 - e^{\frac{t}{a}\ln 3} & t \in [0, a] \\ 2 & t \in a + [0, s+r], \end{cases} \tag{7}$$

and the ratio $f_c/f_m = 2/3$, we synthesized the sound of the clarinet [4], [5]. Figure 2 shows the plots of the clarinet acoustic signal $x(t)$ and its instantaneous frequency $f_i(t)$ for $a = r = 0.15$, $s = 0.7$, $f_c = 20$Hz, and $f_m = 30$Hz (again, the frequency values were chosen for illustration purposes).

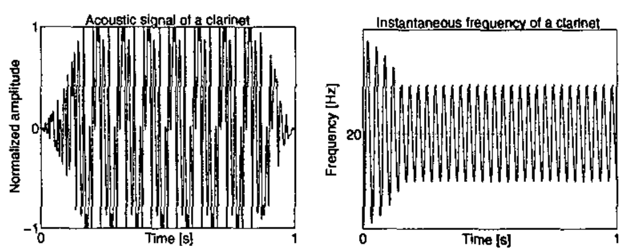

Fig. 2. Clarinet. Left: time evolution of the acoustic signal. Right: its instantaneous frequency.

From the left plot of Figure 2, we see that the time envelope of the acoustic signal $x(t)$ is as defined in (6). To analyze the instantaneous frequency $f_i(t)$, consider two intervals: for $t \in [0, a]$, $I(t)$ is exponential, thus $f_i(t)$ is as in (5), i.e., it is an underdamped oscillation with frequency f_m and mean value f_c; for $t \in [a, a+s+r]$, $I(t)$ is constant, thus, from (4), $f_i(t)$ is sinusoidal (constant amplitude) with mean value f_c and frequency f_m. The right plot in Figure 2 confirms this analysis.

III. Mapping bits into musical notes

To map the digital messages into the musical sounds we used FSK. Standard communications engineering FSK assigns each symbol to a sinusoidal carrier frequency. Our FSK modulation scheme assigns each symbol to a synthesized musical note of a given fundamental frequency. The notes are chosen in such a way that their spectra do not collide. This way our receiver can perceive the transmitted notes just by detecting their fundamental frequency, just like a standard FSK receiver. Besides being computationally very simple, our receiver doesn't even need to know which is the instrument being synthesized by the transmitter (our experiments show that the instrument being synthesized can change whenever wanted without affecting the transmission). The price to pay for this versatility is the low spectral efficiency and the consequent low bit rate.

In the case of the piano, the fundamental frequency of each note is obviously f_0, as defined in (1). For the bells and the clarinet, the spectral analysis of the signal $x(t)$ defined by (2,3) is not trivial. In Figures 3 and 4 we plot the spectrograms of the synthesized sounds of a bell and a clarinet note. The bell sound was obtained with carrier frequency f_c = 500Hz and modulating frequency f_m = 1000Hz. From the spectrogram in Figure 3, we see that the fundamental frequency of the bell sound is f_0 = 500Hz, i.e., it is equal to the mean value f_c of the instantaneous frequency $f_i(t)$, as derived in the previous section. The clarinet sound was obtained with f_c = 2000Hz and f_m = 3000Hz. From the spectrogram in Figure 4, we see that the fundamental frequency of the clarinet note is f_0 = 1000Hz, which is different from the mean value f_c = 2000Hz of the instantaneous frequency. In fact, we show elsewhere that the fundamental frequency f_0 of $x(t)$ in (2,3) is the greatest common factor of the carrier frequencies f_c and the modulating frequency f_m. Thus, according to the definitions of the previous section, we get $f_0 = f_c$ for the bells and $f_0 = f_c/2$ for the clarinet, which agree with the spectrograms of Figures 3 and 4.

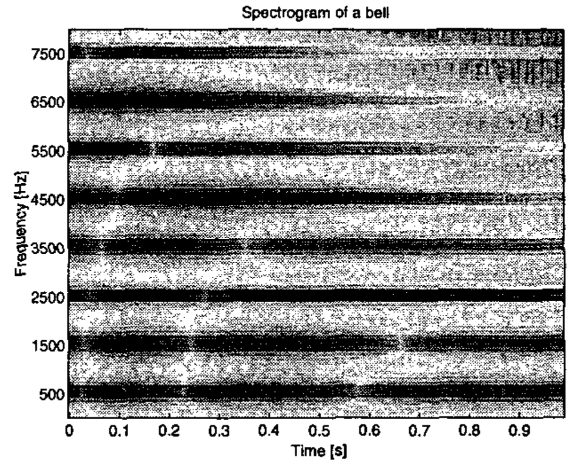

Fig. 3. Spectrogram of a bell sound – carrier frequency f_c = 500Hz, modulating frequency f_m = 1000Hz. The fundamental frequency is f_0 = 500Hz.

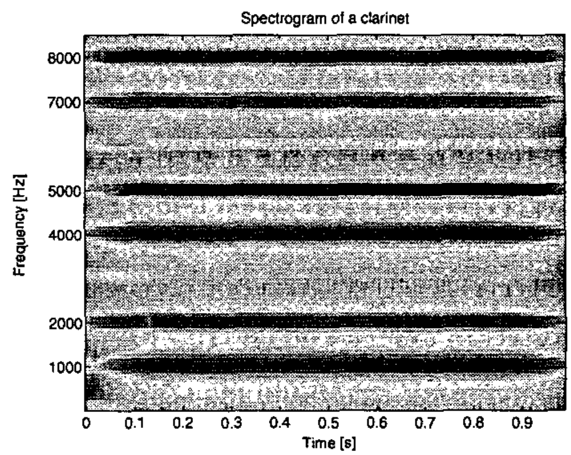

Fig. 4. Spectrogram of a clarinet note – carrier frequency f_c = 2000Hz, modulating frequency f_m = 3000Hz. Note that the fundamental frequency is f_0 = 1000Hz.

We used text files to test our communication system. The coder starts then by transforming the characters of the text file into a binary string using the ASCII code. Then, it groups the bits into symbols. Finally, as just described, the coder assigns each symbol to a fundamental frequency f_0 and produces the corresponding musical note within a time slot of duration T_s. We used two techniques to perform the assignment of the symbols to the frequencies: single channel mapping and multichannel mapping.

Single channel – melody

The single channel bit mapping uses one musical note per time slot T_s, meaning that in each of these time intervals we transmit one symbol. The set of fundamental frequencies selected to represent the symbols has impact on the sound perceived by the humans. We used two sets of frequencies from the ones reference [1] proposed for pure tones: i) equally spaced frequencies; ii) frequencies corresponding to the notes of a chosen key chord. In this last case, we also implemented a *Spread-Spectrum* technique [1] that enables the key chord to change across time, just like what happens in a music melody.

Multichannel – melody and harmony

In contrast with the single channel case, n-channel mapping ($n > 1$) uses n simultaneous notes in each time slot T_s. Thus, in each of these time intervals, we transmit n symbols. We use n mutually exclusive frequency bands so that the waveforms sharing the same time slot would be orthogonal if they were pure tones – *frequency division multiplexing* (FDM). Obviously, broad spectrum signals as our sounds are not made orthogonal just by choosing the fundamental frequency in disjoint bands, so, besides this, we also took into account the main spectral components of each channel sound set. In each of the frequency bands the selection of the frequencies is analogous to the single channel case. With $n = 2$ channels, the resulting acoustic signal contains two musical components that we choose to belong to the same key chord to produce musical harmony, besides melody.

IV. Detection

Our receiver contains basically a standard FSK demodulator. Synchronization is achieved by signaling – each message is preceded and succeeded by a note known to the receiver. We decided to use this signaling scheme because it proved to be much more robust than simply detecting when acoustic power raises. To find the beginning of the message, the receiver continuously correlates the received signal with the sinusoid with frequency equal to the fundamental frequency of the signaling note, over each half time slot interval. Once an half time slot interval is found that corresponds to a correlation above a threshold, the receiver refines the search by correlating over a number of entire time slots that contain that half time interval.

After synchronization is achieved, the receiver infers the transmitted symbol(s) by computing the correlation between the received signal and each of the pure tones in the frequency set used and selecting the one that refers to the correlator with maximum output [6]. Although not in the presence of a pure FSK signal, this incoherent FSK receiver can still be used because the signal's spectral density shows a pronounced maximum in a predefined frequency, see the examples of spectra in Figures 3 and 4.

V. Experiments

To test the communication system, we used ordinary low cost microphone and speaker. Table I summarizes experiments 1 through 12. The distance between the transmitter and the receiver was approximately 2m in experiment 1 and 40cm in the other experiments. We used fundamental frequencies starting at about 400Hz, symbol duration $T_s = 1$s, and coded 4 bits per symbol. In Table I, ES, CH, and SS stand, respectively, for the equally spaced, the chord, and the spread spectrum frequency mappings described in section III. When using ES, the frequency spacing was 20Hz.

Experiments 1 and 11, using pure FSK, provided us a means of comparison with the other system designs. We concluded about the dramatic difference between our musical sounds and simple pure tones. The sounds produced, as well as the source code used, are available in [3]. Experiments 2, 3, and 4 illustrate communications based on each of the instruments synthesized. Experiments 5 and 6 demonstrate the impact of choosing the fundamental frequencies from a musical chord. In experiments 7 and 8, the transmitter uses, respectively, two and three instruments – it swaps from instrument to instrument (for example in experiment 7, every 2s we hear a bell ring). The acoustic signal obtained this way is more music-like. Experiment 9 demonstrated that the SS mapping technique makes the sound less monotonous and more pleasant. Experiment 10 combines SS with the use of the three instruments. Experiment 12 uses two channels with piano sound in both channels.

#	Instrument	# ch.	map	bitrate	errors
1	tuning fork	1	ES	4bps	0%
2	piano	1	ES	4bps	0%
3	bells	1	ES	4bps	3%
4	clarinet	1	ES	4bps	0%
5	piano	1	CH	4bps	0%
6	clarinet	1	CH	4bps	3%
7	clarinet/bells	1	ES	4bps	0%
8	clarinet/piano/bells	1	ES	4bps	0%
9	piano	1	SS	4bps	0%
10	clarinet/piano/bells	1	SS	4bps	0%
11	tuning fork	2	ES	8bps	0%
12	piano	2	ES	8bps	3%

TABLE I

Experiments 1-12. Hear the sounds in the WWW link [3].

To improve the subjective quality of the sound, we made a last experiment using two channels with their own characteristics: channel 1, standard FSK, $T_s = 1.6$s, 4 bits per symbol; channel 2, bells/clarinet, $T_s = 0.8$s, 2 bits per symbol. We got 0% errors with 5bps. The resulting sound has a kind of rhythm, which makes it one of the most music-like messages.

VI. Conclusion

This paper describes an aerial acoustic communication system that produces sounds that are tolerant, or even pleasant, to humans. To transmit a message, the coder synthesizes the sounds of musical instruments such as the piano, the bells or the clarinet and associates each symbol to musical note(s). By carefully choosing the set of musical notes used, the decoder is kept computationally simple. Our experiments show that this system is well suited to applications requiring the transmission of short messages. The first steps to improve the data rate without requiring larger transmitted power or producing higher error probability, would be to use source and channel coding techniques and receivers based on the optimal detectors.

References

[1] C. V. Lopes and P. M. Q. Aguiar. Aerial acoustic communications. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, 2001.

[2] V. Gerasimov and W. Bender. Things that talk: using sound for device-to-device and device-to-human communication. *IBM Systems Journal*, 39(3-4), 2000.

[3] `http://www.isr.ist.utl.pt/~aguiar/code.html`.

[4] J. M. Chowning. The synthesis of complex audio spectra by means of frequency modulation. *Journal of the Audio Engineering Society*, 21(7), 1973.

[5] J. H. McClellan, R. W. Schafer, and M. A. Yoder. *DSPfirst – A Multimedia Approach*. Prentice-Hall, 1999.

[6] B. Sklar. *Digital Communications – Fundamentals and applications*. Prentice-Hall, 1988.