

Real-Time Audio Buffering for Telephone Applications

Paul H. Dietz

Mitsubishi Electric Research Laboratories,
Cambridge Systems
201 Broadway
Cambridge, MA 02139,USA
Tel: 1-617-621-7532
E-mail: dietz@merl.com

William S. Yerazunis

Mitsubishi Electric Research Laboratories,
Cambridge Research
201 Broadway
Cambridge, MA 02139,USA
Tel: 1-617-621-7530
E-mail: yerazunis@merl.com

ABSTRACT

A system that uses an ear proximity sensor to actively manage periods of distraction during telephone conversations is described. We detect when the phone is removed from the ear, record any incoming audio, and play it back when the phone is returned to the ear. By dropping silent intervals and speeding up playback with a pitch-preserving algorithm, we quickly return to real-time without the loss of information. This real-time audio buffering technique also allows us to create a user-activated, lossless instant replay function.

KEYWORDS: Real-time audio buffering, instant replay

INTRODUCTION

There are brief moments when people are distracted from their telephone conversations. Examples include interruptions from coworkers, baby screams, gear shifting, or daydreaming. In many of these cases, a user purposefully removes the phone from his or her ear to address the other task. In other cases, the user was trying to hear, but failed.

In this work, we present a system that uses an ear proximity sensor to actively manage these periods of distraction. We detect when the phone is removed from the ear, record any incoming audio, and play it back when the phone is returned to the ear. By dropping silent intervals and speeding up playback with a pitch-preserving algorithm, we quickly return to real time without loss of information. We call this technique *real-time audio buffering*.

A potential problem with such a system is that a user loses the context of the conversation while distracted. For this reason, the playback includes several seconds of content from before the phone was removed from the ear.

When a user fails to hear what is said, a lossless instant replay function is very useful. This functionality is a side effect of the system already described. The user briefly

moves the phone away and then back to the ear. This causes the playback of material from a short time ago. Each time the handset is moved away and back, playback resumes successively further back in time, allowing the user to replay arbitrarily long segments, within memory limits.

The concept of accelerated playback is well known in the literature (for example, [2, 4]) and has been used in equipment such as dictation machines for many decades. Similarly, the basic idea of catching up to real-time despite disruptions has been described in [1]. The major contribution of this work is the unique user interface.

In the following sections, we examine some implementation details, and discuss our results with a prototype system.



Figure 1: the ear-sensing handset.

REAL-TIME AUDIO BUFFERING

Real-time audio buffering cleanly transitions from time-compressed audio to real time audio. Incoming audio is stored in a circular buffer at a constant rate. In normal use, the playback pointer simply tracks the record pointer. When the phone is removed from the ear, the playback pointer is pushed back several seconds, and held constant. When the phone is returned to the ear, playback resumes. While the playback pointer is trailing the record pointer, the playback is sped up by dropping silent segments and using a simple pitch-preserving time-compression algorithm. This continues until the playback pointer catches up to the record pointer and normal playback is resumed. Since the buffer is circular, holding the phone away from the ear for an extended period of time will cause the record pointer to “lap” the stalled playback pointer. In this case, we have no choice but to move the playback pointer so as to track the oldest, non-overwritten data.

TEST HARDWARE

Our test system consists of a PC with a sound card and two handsets (Riparius INT100 Internet Telephone Handsets). One handset is modified to provide non-contact capacitive proximity sensing [3], which detects ear presence.

The modified handset is shown in figures 1-3. Electro-conductive paint (MG Chemicals "SuperShield") is applied to the interior of the handset to create a large ground electrode, and a separate, small, isolated ear electrode. Wires are attached with conductive epoxy (CircuitWorks CW2400). In use, the user's hand capacitively couples the large ground plane to the ear electrode via the user's ear. In our prototype, this capacitive coupling is detected by a small circuit that reports these measurements to the PC via an RS-232 serial link. For convenience, the circuit is parasitically powered from this RS-232 port. This circuit is tiny, reliable, draws little power, and can be implemented in current cell phones at minimal cost.

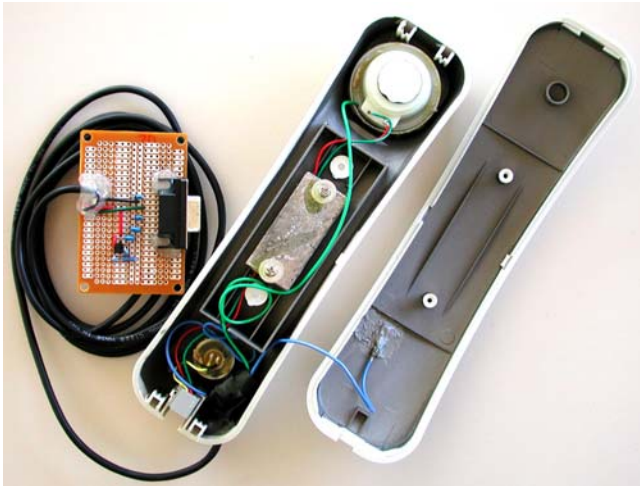


Figure 2: inside the ear-sensing handset.



Figure 3: the ear electrode (speaker removed).

SIGNAL PROCESSING

All signal processing is implemented on a PC running Red

Hat Linux version 7.1. Our time-compression algorithm is based on the Fairbanks technique [1] - we first remove long periods of silence, and then we drop every Nth segment (each segment is 256 samples at 16000 samples/sec). While better sounding algorithms exist [4], the Fairbanks algorithm is easy to implement, requires few computational resources, and is adequate for our needs

RESULTS

The prototype implementation has been ad hoc tested with about 20 users. Training needed to use the system is minimal. It suffices to say "this phone will pause whenever you take it away from your ear, and resume automatically when you start listening again." Counting from one to twenty, while the users alternate listening and pausing, confirms to them that the system performs as expected. No user required more than sixty seconds of training. The majority of testers thought this to be an excellent feature that should routinely be included on telephones.

FUTURE WORK

Since cell phones are the target of this work, we hope to leverage the calculations performed by the phone's codec to create more intelligible algorithms for speech acceleration. Once this is completed, extensive user studies are needed to refine the algorithm and measure the efficacy of the technique.

CONCLUSIONS

We have demonstrated how real-time audio buffering is an effective means of recovering from brief distractions while on the telephone. The use of an ear sensor provides a mostly automatic, and extremely intuitive user interface.

ACKNOWLEDGMENTS

The authors would like to thank Michael Casey for his assistance with time-compressed speech algorithms and Darren Leigh for his assistance getting various bits and pieces of the system working. Also, we would like to thank the reviewers for their insightful comments.

REFERENCES

1. Ahmed, S. et al, Display Pause with Elastic Playback, U.S. Patent 6,005,564, 1999.
2. Arons, B. Techniques, Perception, and Applications of Time-Compressed Speech. In *Proceedings of 1992 Conference, American Voice I/O Society*, 1992.
3. Baxter, L. *Capacitive Sensors Design and Applications*. IEEE Press, Piscataway, NJ, 1997.
4. Covell, M., Withgott, M, and Slaney, M., Mach1: Nonuniform Time-Scale Modification of Speech. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 1998.