

Humming Control Interface for Hand-held Devices

Sook Young Won
Stanford University
660 Lomita Drive
Stanford, CA, U.S.A
sywon@ccrma.stanford.edu

Dong-In Lee
Stanford University
660 Lomita Drive
Stanford, CA, U.S.A
joshua79@ccrma.stanford.edu

Julius Smith
Stanford University
660 Lomita Drive
Stanford, CA, U.S.A
jos@ccrma.stanford.edu

ABSTRACT

This paper describes a control-by-humming interface in which a bluetooth-connected insertion earphone/microphone remotely controls a small portable system such as a modern assistive device, cell phone, etc. A pitch detection algorithm converts a subvocal hum input signal into pitch contours that are segmented into discrete “notes” and then grouped to form control commands. These commands cause transitions among operational states. An example application is given for hands-free control of a simplified (six-state) cell phone and music player system. Performance of the interface is discussed and future improvements are outlined.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces—*Input devices and strategies (e.g., mouse, touch-screen), Voice I/O*

General Terms

Design, Human Factors

1. INTRODUCTION

To maximize the utility of alternative device controllers, user requirements must be minimized. One class of alternative controllers is hands-free, voice-operated control. Often a subset of natural language is used for this purpose. On the other hand, controllers using non-verbal sounds, such as singing, whistling, and humming have also been explored, such as the acoustic mouse [6] and vocal joystick [1]. In this paper, we summarize our recent work on the use of *subvocal humming* as a means of control.

Humming minimizes user requirements because it is significantly easier to produce than speech, being possible even without vocal folds. Compared with natural language processing, hum processing requires much less computation and power, and is independent of language and accent. More specifically, *subvocal* humming is uniquely unobtrusive (to

other audience members in a movie theater, for example) when compared with traditional voice-based or whistle-based control. This natural unobtrusiveness is maximized by the use of an internal sensor such as a contact microphone inserted in the outer ear to receive bone-conducted sound. Another major advantage of bone-conducted sound input is immunity to external environmental noise.

The control-by-humming design discussed in this paper is further influenced by the ongoing convergence of programmable cell phones, music players, and assisted-listening devices. Examples of such systems include the Nokia N95 and Apple iPhone. Having both input and output audio opens up a larger space of control strategies, e.g., using audio “menus” and other kinds of user-directed information retrieval.

In our design, we assume the availability of a wireless “bluetooth earphone” which fully inserts into one ear (optionally using two for stereo)[5]. We also assume the earphone to have a bone-conduction microphone for picking up subvocal humming sounds, and a conventional air-pressure microphone in the ear canal, for picking up environmental sounds.

The remainder of this paper presents our current control-by-humming interface for integrated personal audio management. While we do not, in this paper, explicitly propose a specific control application for the motor-impaired, we believe the ideas and methodology are readily applicable in this domain. Our overall goal is a comprehensive personal audio and device management system controlled unobtrusively by subvocal humming in a hands-free manner. Ideally, a user will be able to go all day without needing to remove his or her earphone-microphones. The present paper is a first installment to document our input signal processing, control vocabulary formation, and state machine management.

2. HUM PROCESSING

There are four stages in converting the voice signal to a message:

1) *Pitch Detection* - For estimating pitch from an input signal, we examined the ‘YIN’ algorithm [3] which is based on the well known autocorrelation method for detecting a signal’s periodicity.

2) *Note Segmentation* - Next, the system distinguishes meaningful notes based on the slope of neighboring pitches and duration of sections. Thus, if the pitch contour of a section is relatively flat, and the length of the section is longer than a designated time threshold, the system regards the section as a note.

3) *Message Generating* - We assign the strings ‘B’, ‘Up’ and ‘Dwn’ to the first note, rising contour, and falling contour of neighboring two notes, respectively. After completing an input query in a string format, we adopt a fast and noise-robust algorithm called the *Levenshtein distance* [4] for string matching.

4) *State Transition* - The transition is decided by two or three values of a current state, an array of previous states, and the generated message or interruption.

3. EXAMPLE APPLICATION

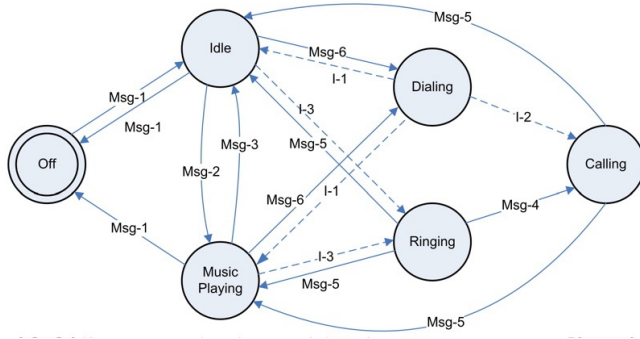


Figure 1: State transitions by humming messages and external interruptions.

Figure 1 depicts the state transition diagram of a simple six-state example (an integrated cell phone and an music player). Solid lines marked “Msg-1” to “Msg-6”, and dashed lines marked “I-1” to “I-3” represent *active* transition messages by the user and *passive* interruptions from the outside, respectively. The messages (device features) and the pitch sequence for control-by-humming for each message are described in Table 1. ¹Three interruptions, I-1 to I-3, happens: 1)When the callee does not answer the user’s call, 2)When the callee answers the call, and 3) when someone calls the user on the appropriate states, respectively.

Table 1: Humming controls generating the active message

	Messages	Pitch Contour	Example
1	Headset On / Off	one long note	Do
2	Music On	three raising notes	Do Re Mi
3	Music Off	three falling notes	Mi Re Do
4	Answering Call	two raising notes	Do Mi
5	End./Passing Call	two falling notes	Mi Do

4. PERFORMANCE EXAMPLE

The following figure is an example use of the implemented system described above, which starts at the initial state ‘Headset off’, goes through several states by messages and interruptions, then comes back to the first state.

¹“Msg-6” for *Dialing(Making Calls)* includes several methods for initiating calls by humming such as to encode speed-dial numbers or to navigate and select from a list of most recently dialed numbers, or from stored numbers, etc.

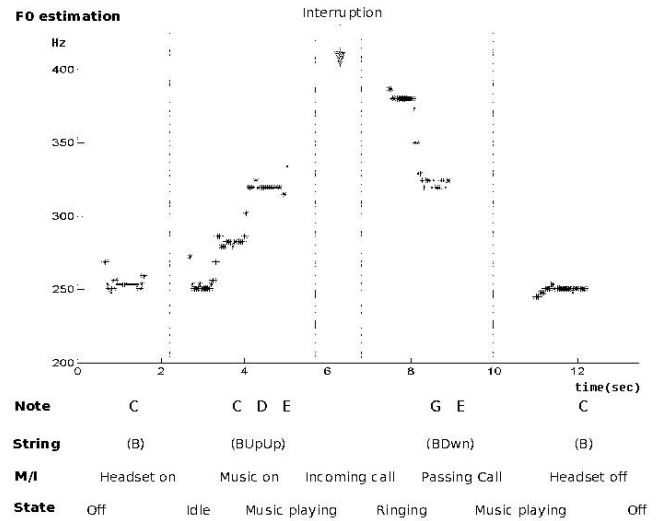


Figure 2: Example use of the system involving a sequence of four steps.

5. CONCLUSIONS AND FUTURE WORK

The current imlementation successfully segments a hum input signal based on pitch, and classifies segments as discrete commands. The YIN pitch-detection and the Levenshtein string-matching algorithms are performing satisfiorly in this application. The state-machine example integrating basic cell-phone and music-player commands is ready for added functionality such as motor-control commands and assisted listening features. Moreover, since all the needed components are in place, we are considering integrating a query-by-humming system [2], which finds music based on a hummed excerpt.

6. REFERENCES

- [1] J. A. Bilmes and et al. The vocal joystick: A voice-based human-computer interface for individuals with motor impairments. In *Proc. Human Language Technology Conference*, Vancouver, Canada, 2005.
- [2] R. B. Dannenberg et al. A comparative evaluation of search techniques for query-by-humming using the musart testbed. *J. Amer. Soc. Info. Science and Tech.*, 58(3):237–245, February 2007.
- [3] A. de Cheveigné and H. Kawahara. YIN, a fundamental frequency estimator for speech and music. *J. Acoust. Soc. America*, 111(4):1917–1930, April 2002.
- [4] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. soviet physics doklady. *Soviet Physics Doklady*, 10:707–710, 1966.
- [5] Z. Liu, M. L. Seltzer, A. Acero, I. Tashev, Z. Zhang, and M. Sinclair. A compact multi-sensor headset for hands-free communication. In *proceedings of the IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, October 2005.
- [6] A. J. Sporka, S. H. Kurniawan, and P. SlavŠk. Acoustic control of mouse pointer. *Universal Access in the Information Society*, 4(3):237–245, March 2006.