

# Voice-driven Interaction:

## Harnessing the capacity of human voice for controlling computer interfaces

Susumu Harada

Computer Science and Engineering

Jonathan Malkin

Electrical Engineering

Jacob O. Wobbrock

The Information School

Xiao Li

Electrical Engineering (now at MSR)

James A. Landay

Computer Science and Engineering

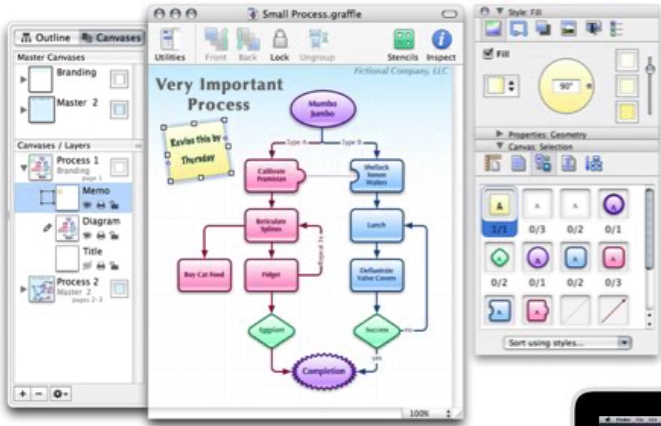
Jeff A. Bilmes

Electrical Engineering

**dub**  
design:  
use:  
build:  
university of washington



UNIVERSITY OF WASHINGTON  
COLLEGE of ENGINEERING  
*Electrical Engineering*



WIKIPEDIA  
*Enciclopedia liberă*







250,000

People with spinal cord injury

47%

are quadriplegic

70%

are unemployed

46,000,000

Adults diagnosed with arthritis

1,000,000

Adults with Parkinson's disease

50,000

Children and adults with  
Muscular dystrophy

Sources:

<http://www.sci-info-pages.com/facts.html>

<http://unitedspinal.org/pdf/scd%20fact%20sheet.pdf>

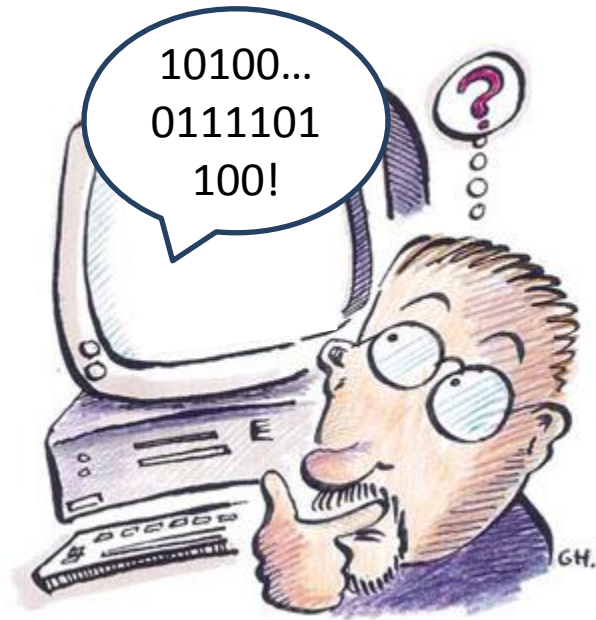
<http://www.hmc.psu.edu/healthinfo/>

# Voice Input Output

Lights, Camera, Action



[http://www.microsoft.com/enable/aging/scan3\\_large.aspx](http://www.microsoft.com/enable/aging/scan3_large.aspx)

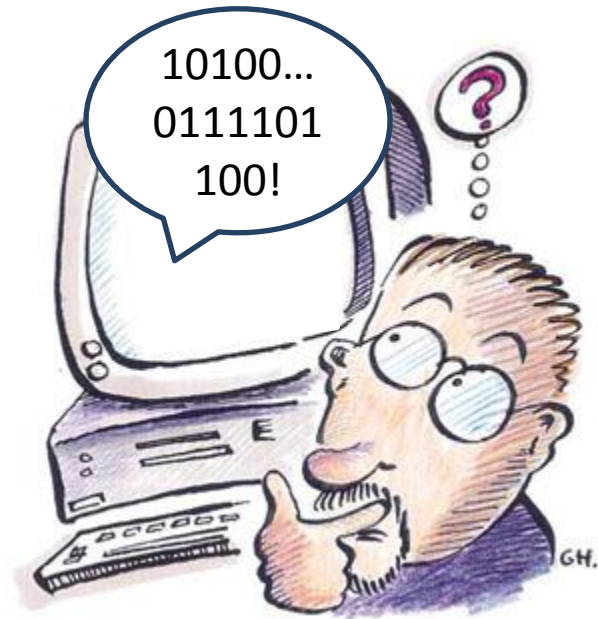


<http://www.ala.org/ala/online/resources/selectedarticles/10reasonswhy.cfm>

# XoIdiotie

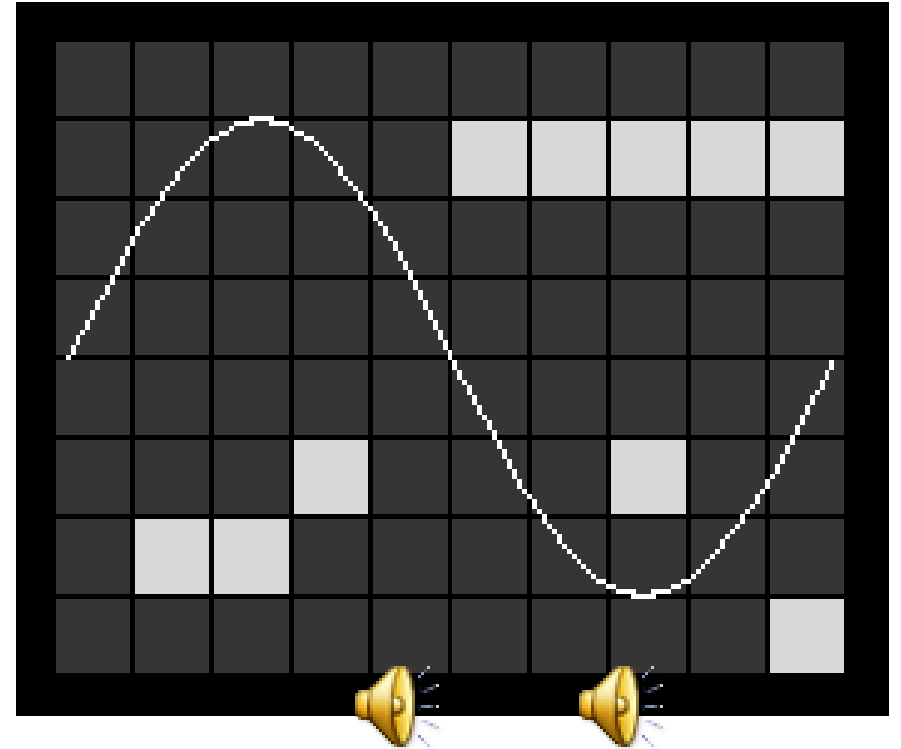
# Output

- **The vOICe**
- **PLUMB**
- **3D Audio Web Browser**
- **Games for the Blind**



# The vOICe

Peter B.L. Meijer. <http://www.artificialvision.com/>

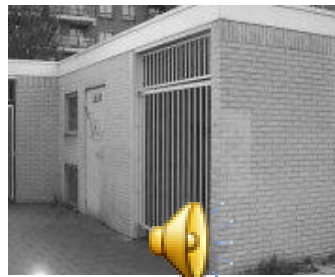
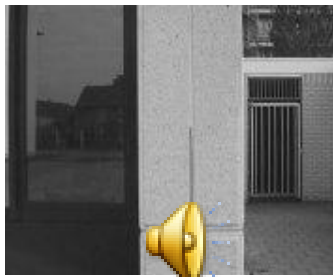


Photography: courtesy Barbara Schweizer



# The vOIce

Peter B.L. Meijer. <http://www.artificialvision.com/>

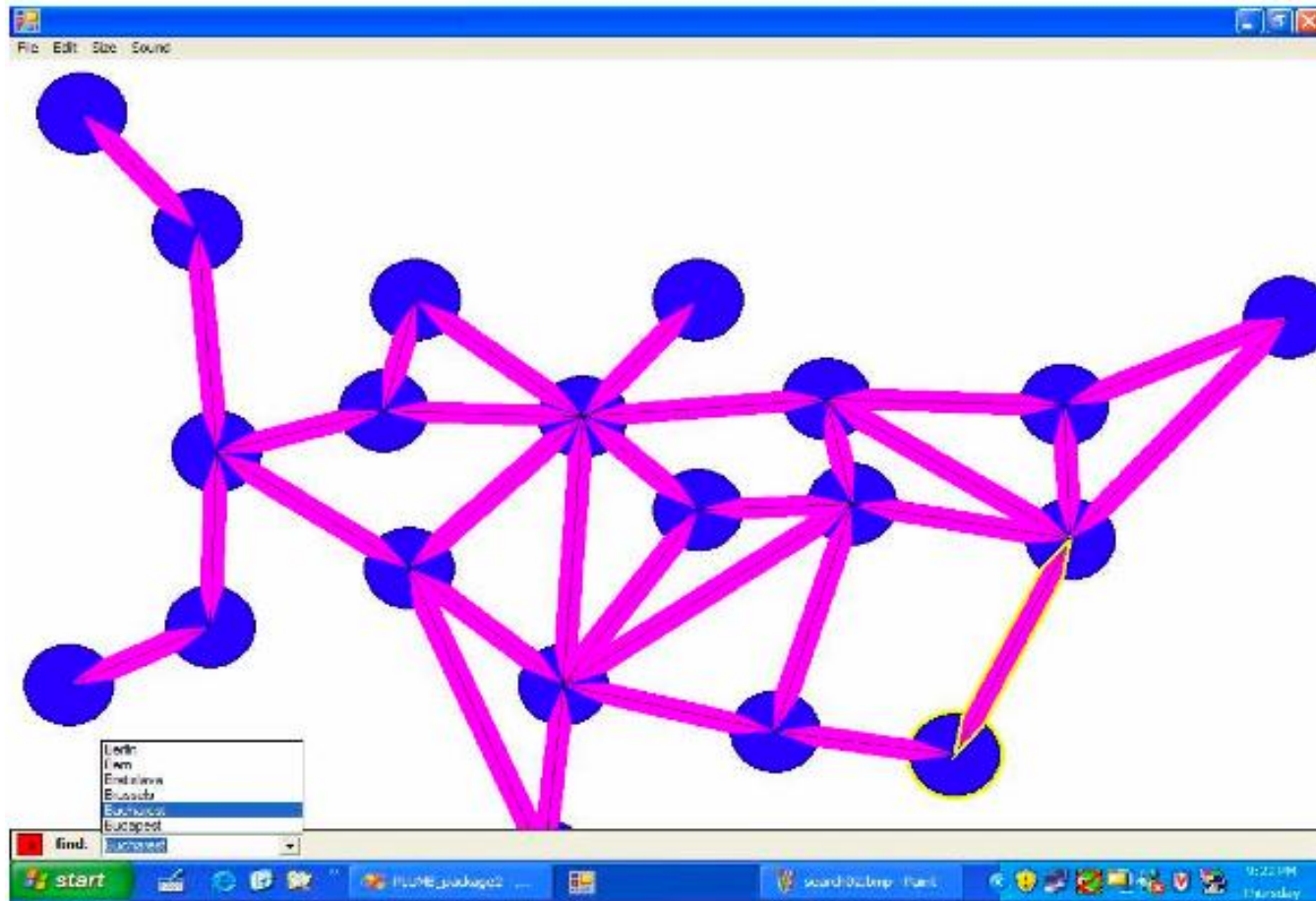


# PLUMB

Matt Calder, Robert F. Cohen, Jessica Lanzoni, and Yun Xu.

*PLUMB: an interface for users who are blind to display, create, and modify graphs.*

ASSETS 2006

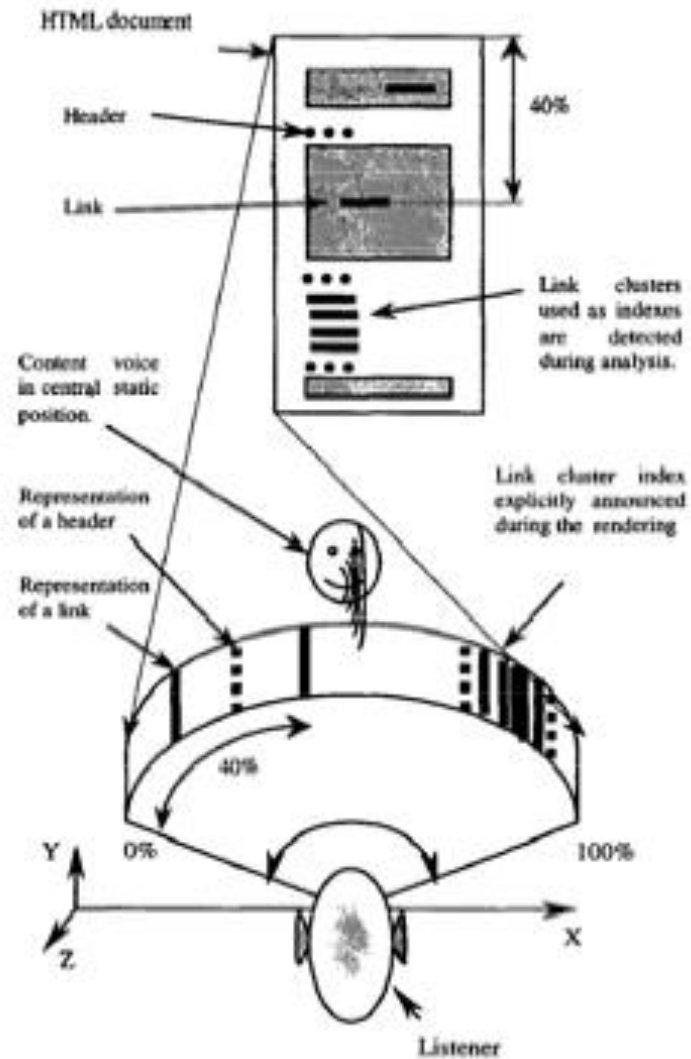


# 3D audio

Stuart Goose and Carsten Möller.

*A 3D audio only interactive Web browser: using spatialization to convey hypermedia document structure.*

MULTIMEDIA 1999



# Shades of Doom

GMA Games.

<http://www.gmagames.com/sod.html>

SHADES OF DOOM 1.2 - Demonstration Mode

**SHADES OF DOOM 1.2**

**A GMA Game**

# Input

Lights, Camera, Action

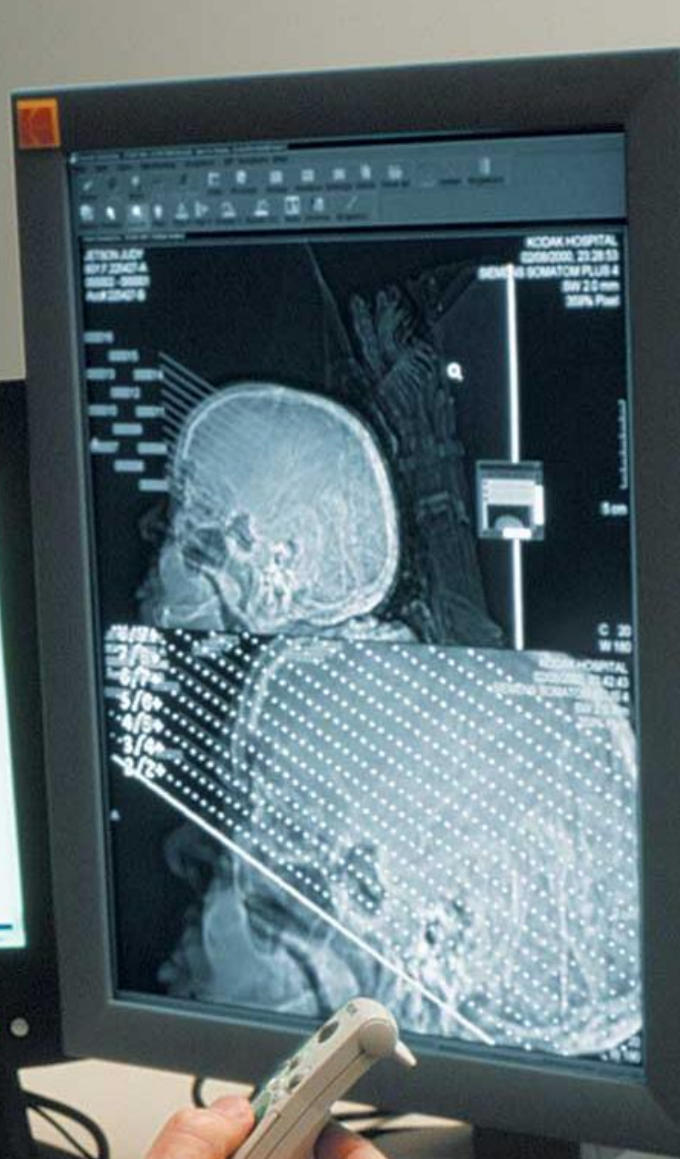


[http://www.microsoft.com/enable/aging/scan3\\_large.aspx](http://www.microsoft.com/enable/aging/scan3_large.aspx)

- **Filled Pause Detection**
- **Pitch-Driven Mode Change**
- **Voice Drummer**
- **Voice as Sound**
- **Migratory Cursor**
- **Whistling UI**
- **Vocal Joystick**
- **VoiceDraw**



Computer, open file.





Arial 10 Western **B** / U



|



# Filled pause

Masataka Goto, Katunobu Itou, and Satoru Hayamizu.

*A Real-time Filled Pause Detection System for Spontaneous Speech Recognition.*

Eurospeech 1999

*A Real-time system*

**Detecting Filled Pauses**

*in Spontaneous Speech*

**1999 Masataka Goto**  
**(goto@etl.go.jp)**

# Speech completion

Masataka Goto, Katunobu Itou, and Satoru Hayamizu.

*Speech Completion: On-demand Completion Assistance Using Filled Pauses for Speech Input Interfaces.*

ICSLP 2002

## Speech Completion

On-demand Completion Assistance

Masataka Goto

*(m.goto@aist.go.jp)*

# Speech spotter

Masataka Goto, Koji Kitayama, Katunobu Itou, and Tetsunori Kobayashi

*Speech Spotter: On-demand Speech Recognition in Human-Human Conversation on the Telephone or in Face-to-Face Situations.*

ICSLP 2004

## **On-Demand Information System for Assisting Human- Human Conversation**

### **Application of Speech Spotter**

Masataka Goto, Koji Kitayama,  
Katunobu Itou, Tetsunori Kobayashi

# Voice drummer

Tomoyasu Nakano, Masataka Goto, Jun Ogata, and Yuzuru Hiraga.

*Voice Drummer: A Music Notation Interface of Drum Sounds Using Voice Percussion Input.*

UIST 2005

## **Voice Drummer**

---

**A Music Notation Interface  
of Drum Sounds  
Using Voice Percussion Input**

by Tomoyasu Nakano  
Masataka Goto  
Jun Ogata  
Yuzuru Hiraga

# Voice as sound

Takeo Igarashi and John F. Hughes.

*Voice as Sound: Using Non-verbal Voice Input for Interactive Control.*

UIST 2001

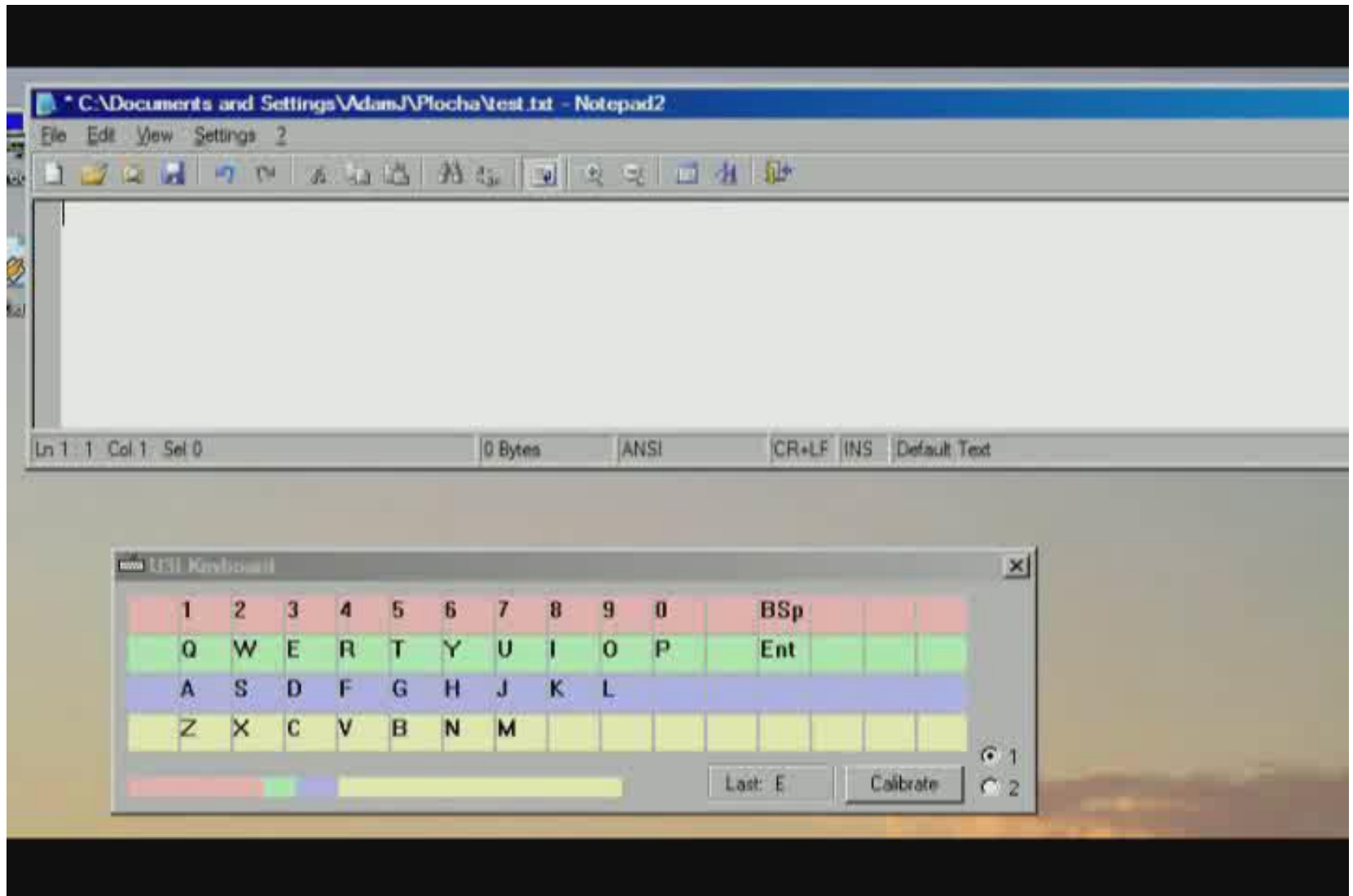
## Voice as Sound: Using Non-verbal Voice Input for Interactive Control

Takeo Igarashi  
John F. Hughes  
(Brown University)



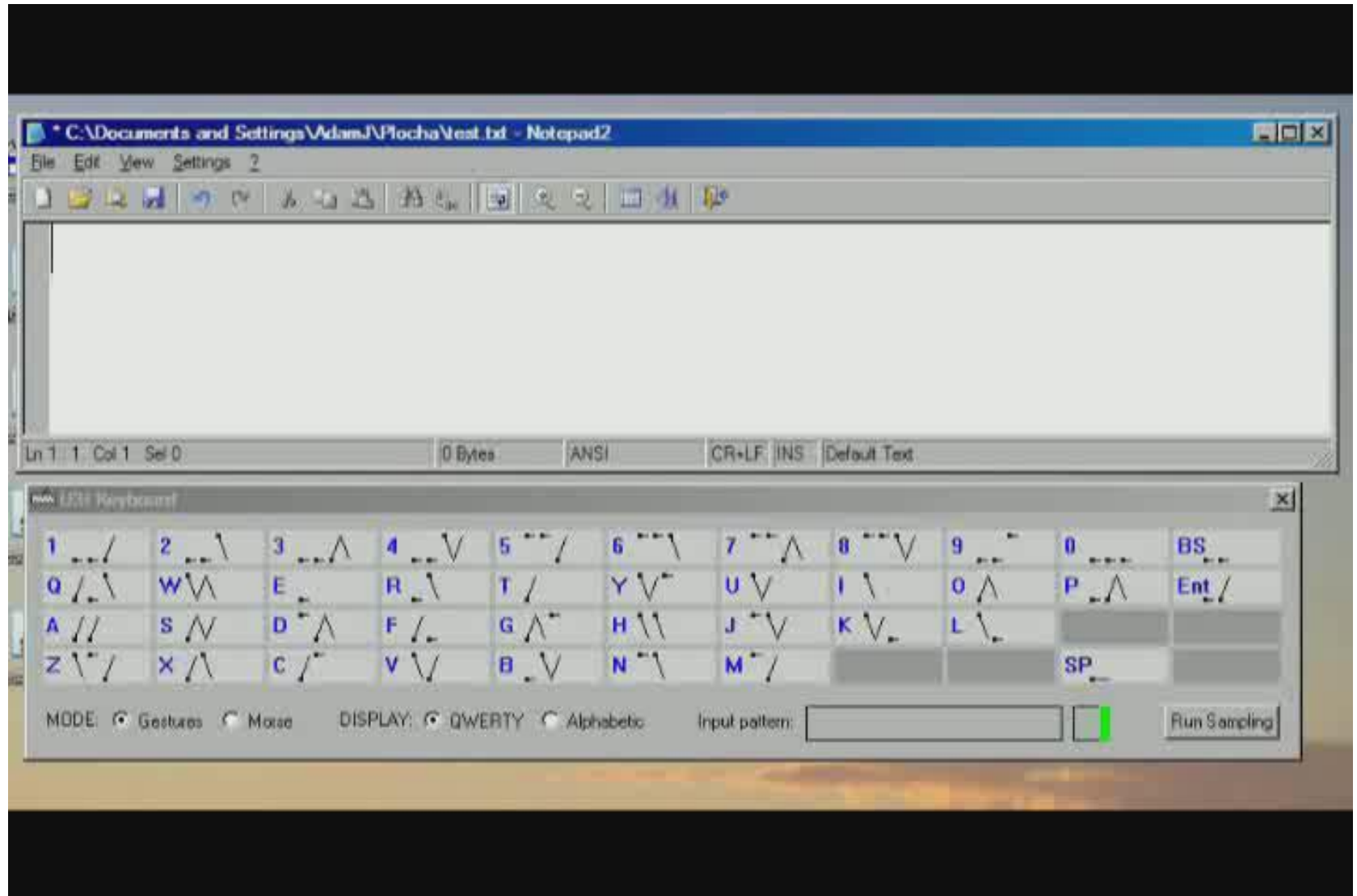
# Whistle text entry

Adam Sporka, Sri Kurniawan, and Pavel Slavik.  
*Non-Speech Operated Emulation of Keyboard.*  
Designing Accessible Technology, 2006



# Whistle text entry

Adam Sporka, Sri Kurniawan, and Pavel Slavik.  
*Non-Speech Operated Emulation of Keyboard.*  
Designing Accessible Technology, 2006

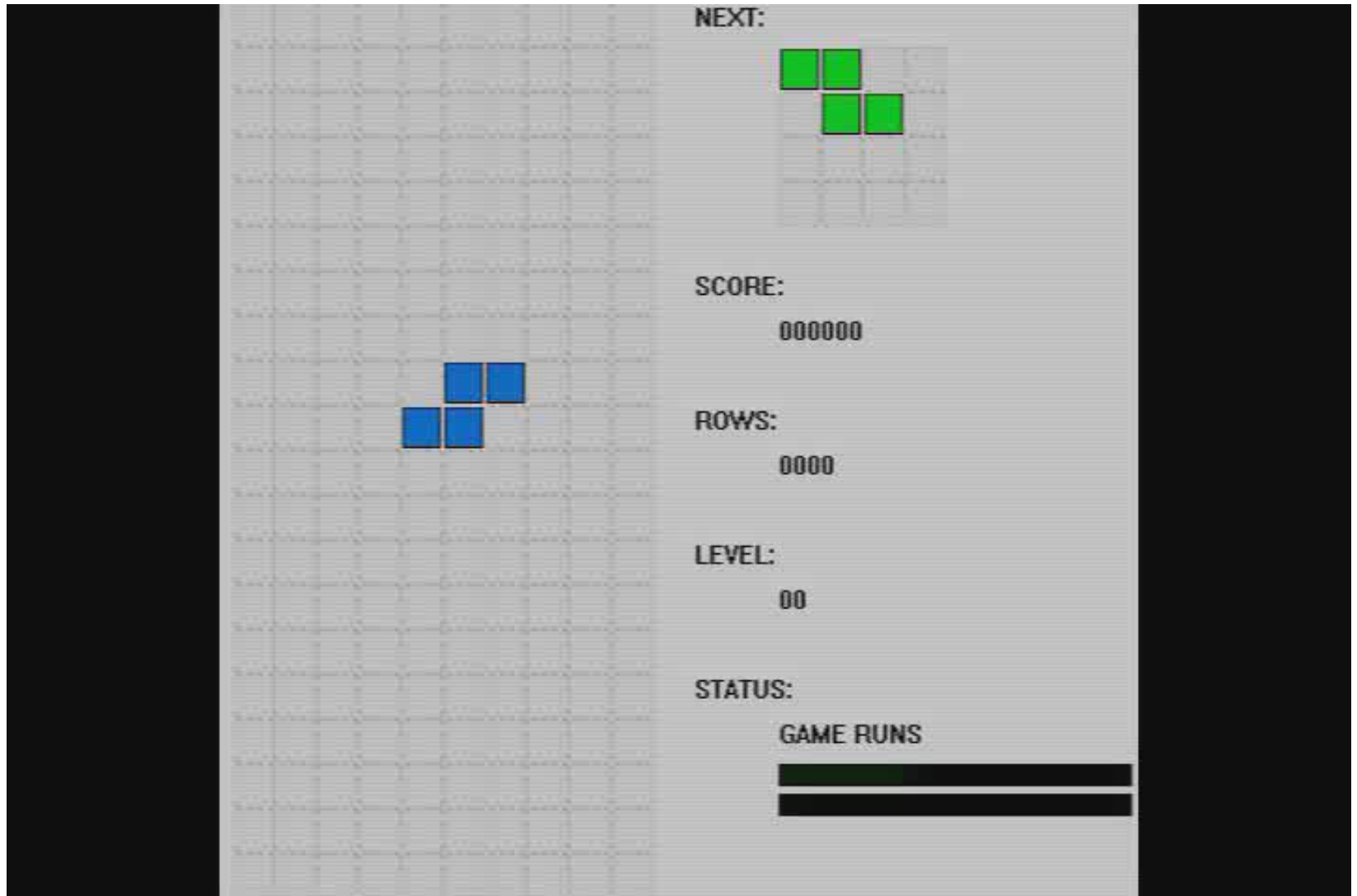


# Humming game

Adam Sporka, Sri Kurniawan, Muni Mahmud, and Pavel Slavik.

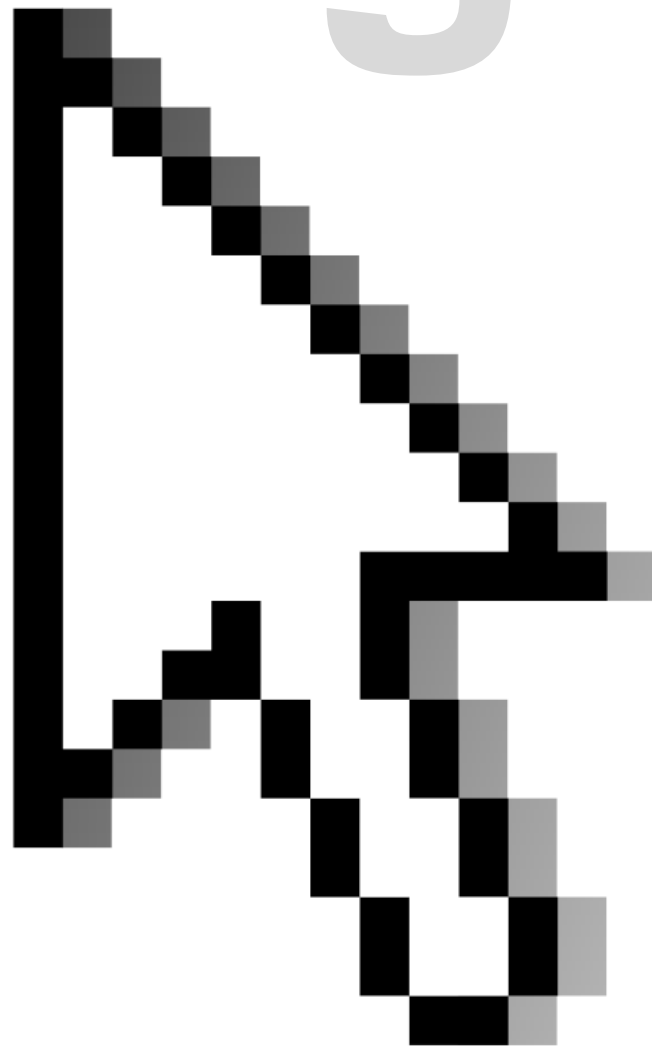
*Non-speech Input vs Speech Recognition: Real-time Control of Computer Games.*

ASSETS, 2006





Pointing  
with  
Voice



# Dragon Dictate

ASSETS 2007 Conference :: October 14-17, 2007 Tempe, AZ, USA - Mozilla Firefox

File Edit View Go Bookmarks Tools Help del.icio.us

http://www.acm.org/sigaccess/assets07/

## ASSETS 2007

October 14-17, 2007 : Tempe, AZ USA



[Home](#) | [Contact](#) | [Conference Organizers](#) | [Program Committee](#) | [Sponsors](#)

### Ninth International ACM SIGACCESS Conference on Computers and Accessibility

[Call For Participation](#)  
[Best Paper Awards](#)  
[Download CFP](#)  
[Doctoral Consortium](#)  
[Student Research Competition](#)  
[Accessibility](#)

[SIGACCESS Homepage](#)

[XML](#)



### Ninth International ACM SIGACCESS Conference on Computers and Accessibility

Computer and Information Technologies have redesigned the way modern society operates. In particular, they have identified new avenues to assist individuals with special needs and provided tools and resources to alleviate the traditional barriers encountered by persons with disabilities. For example, speech generation systems have assisted persons with visual disabilities, voice recognition has helped people with motor impairments, multi-modal presentations have been shown to be

#### Events and Timeline

Paper submissions	25
	May 2007
Acceptance notification for papers	6
	July 2007
Panel, poster, and demonstrations submissions	22
	June 2007
Doctoral Consortium Proposal submissions	13
	July 2007
Student research competition submission	13
	July 2007
Acceptance notification for other venues	20
	July 2007
Camera-ready	8

<http://www.acm.org/sigaccess/assets07/studentresearch/index.php>

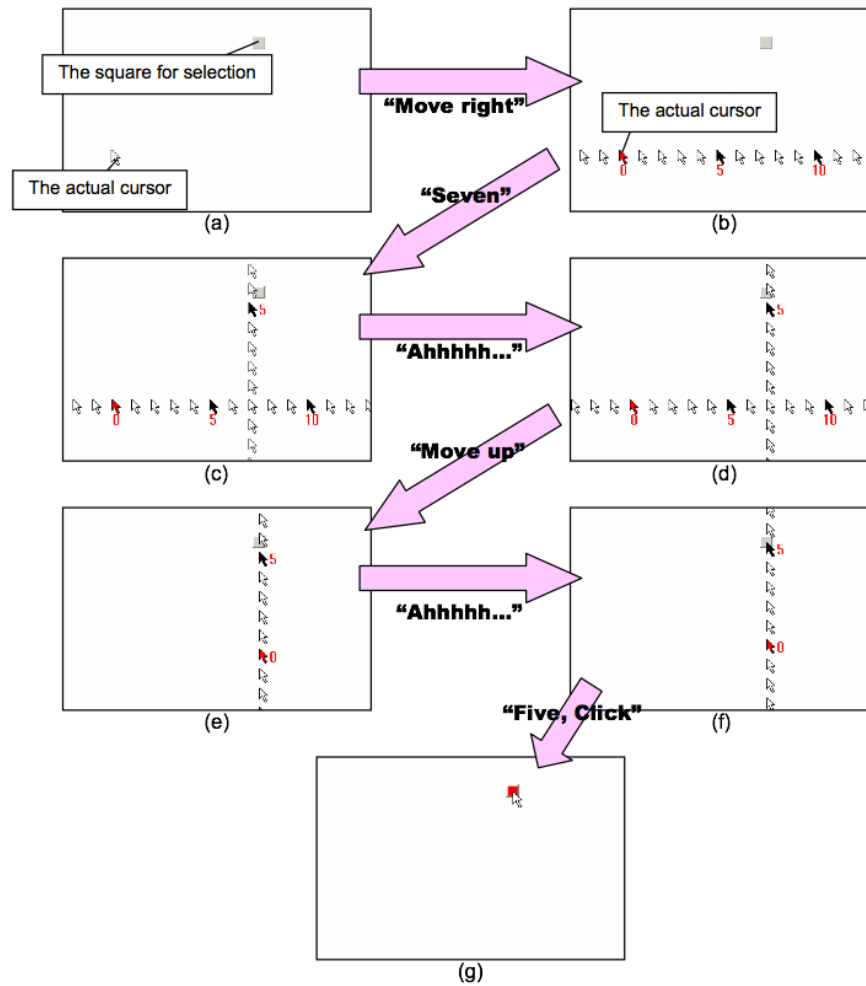
Open Notebook

# Migratory cursor

Yoshiyuki Mihara, Etsuya Shibayama, and Shin Takahashi.

*The Migratory Cursor: accurate speech-based cursor movement by moving multiple ghost cursors using non-verbal vocalizations.*

ASSETS 2005

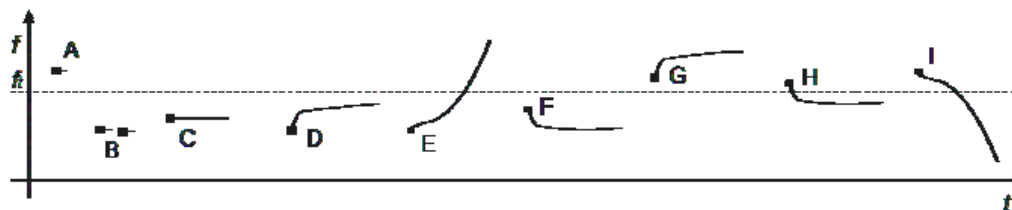
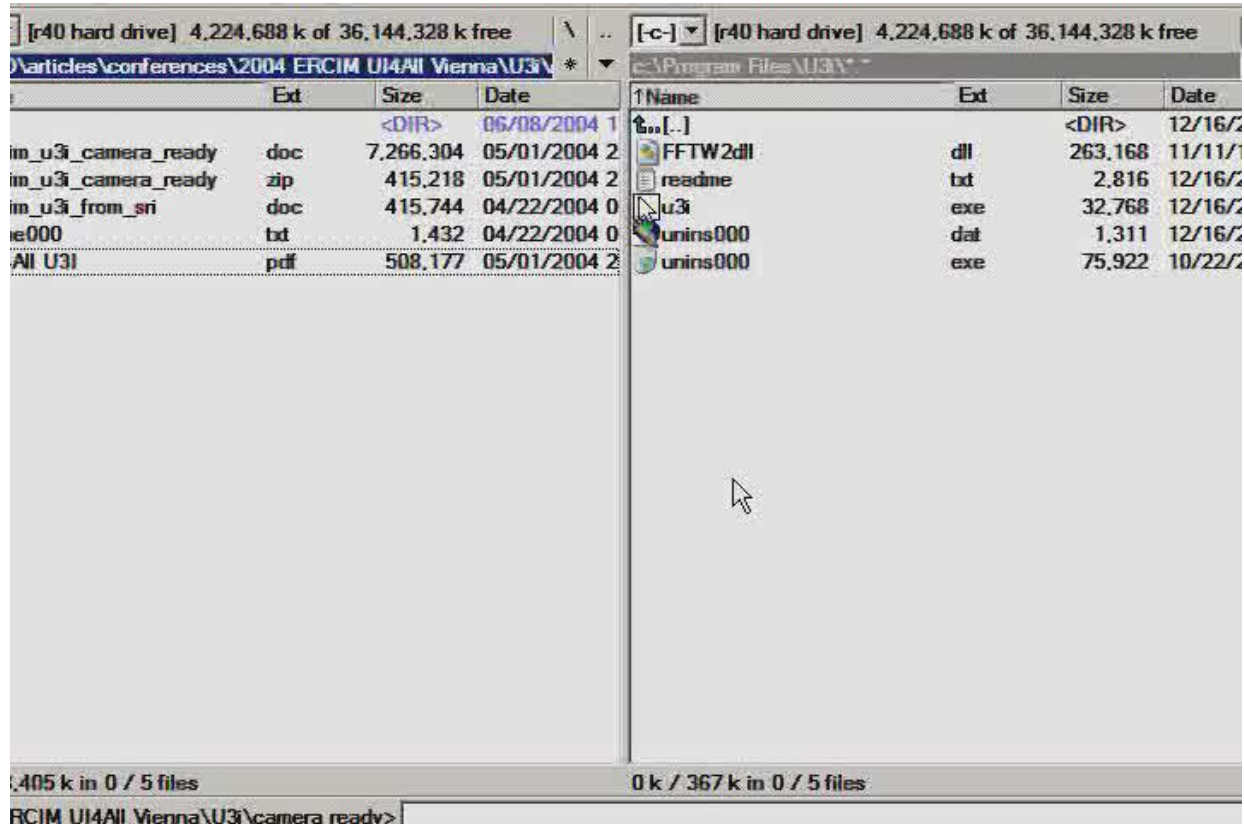


# Whistling UI

Adam Sporka, Sri Kurniawan, and Pavel Slavik.

Whistling User Interface (U3I).

ERCIM 2004



**Fig. 2.** Examples of control tones.  $t$  – time,  $f$  – pitch,  $f_t$  – threshold pitch; A – click, B – double click, C – no motion, D – motion to the right, E – fast motion to the right, F – motion to the left, G – motion up, H – motion down, I – fast motion down.

# Vocal Joystick

The screenshot shows a Mozilla Firefox browser window with the title "assets 2007 - Google Search - Mozilla Firefox". The address bar contains the URL "http://www.google.com/search?q=assets+2007&start=0&". The search bar contains the text "assets 2007" and a "Search" button. The page displays search results for "assets 2007", showing "Results 1 - 10 of about 17,500,000 for assets 2007. (0.20 seconds)".

**Web** Results 1 - 10 of about 17,500,000 for [assets](#) 2007. (0.20 seconds)

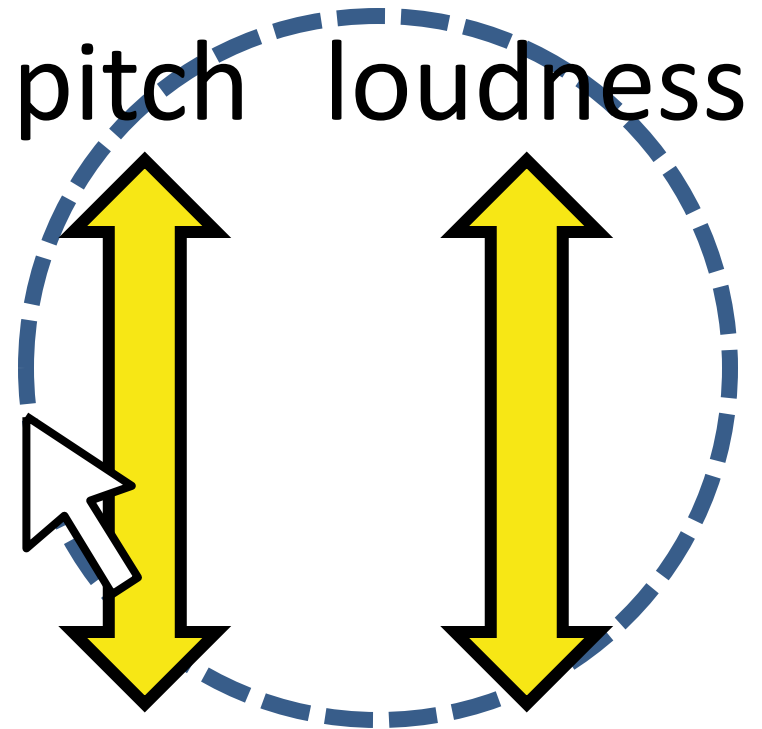
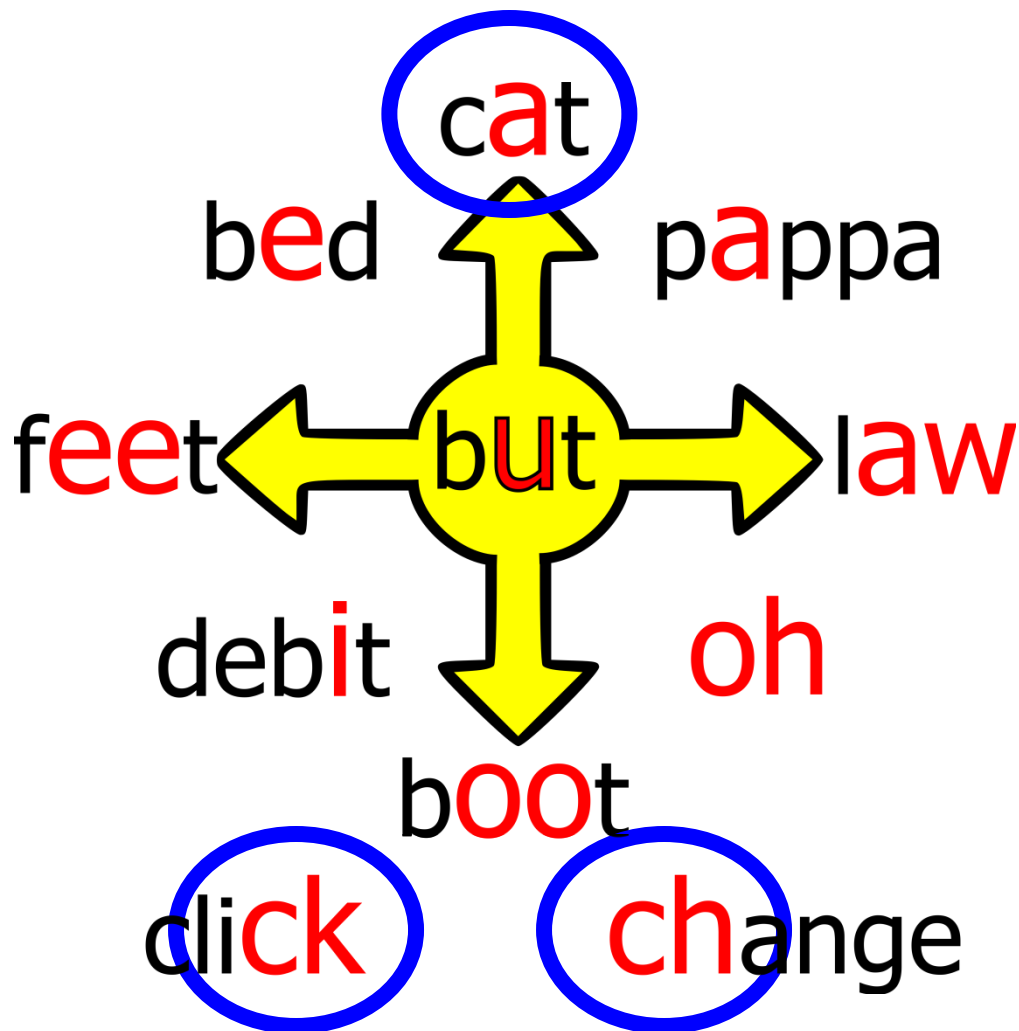
[ASSETS 2007 Conference :: Call For Papers](#)  
ASSETS 2007 9th ACM Conference on Computers and Assessibility.  
[www.acm.org/sigaccess/assets07/cfp/](http://www.acm.org/sigaccess/assets07/cfp/) - 18k - [Cached](#) - [Similar pages](#) - [Note this](#)

[PDF] [ASSETS 2007 The Ninth International ACM SIGACCESS Conference on ...](#)  
File Format: PDF/Adobe Acrobat - [View as HTML](#)  
The ASSETS 2007 Doctoral Consortium provides an opportunity for doctoral students to explore their research interests in an interdisciplinary workshop, ...  
[www.acm.org/sigaccess/assets07/cfp/2007\\_Assets\\_cfp.pdf](http://www.acm.org/sigaccess/assets07/cfp/2007_Assets_cfp.pdf) - [Similar pages](#) - [Note this](#)  
[ [More results from www.acm.org](#) ]

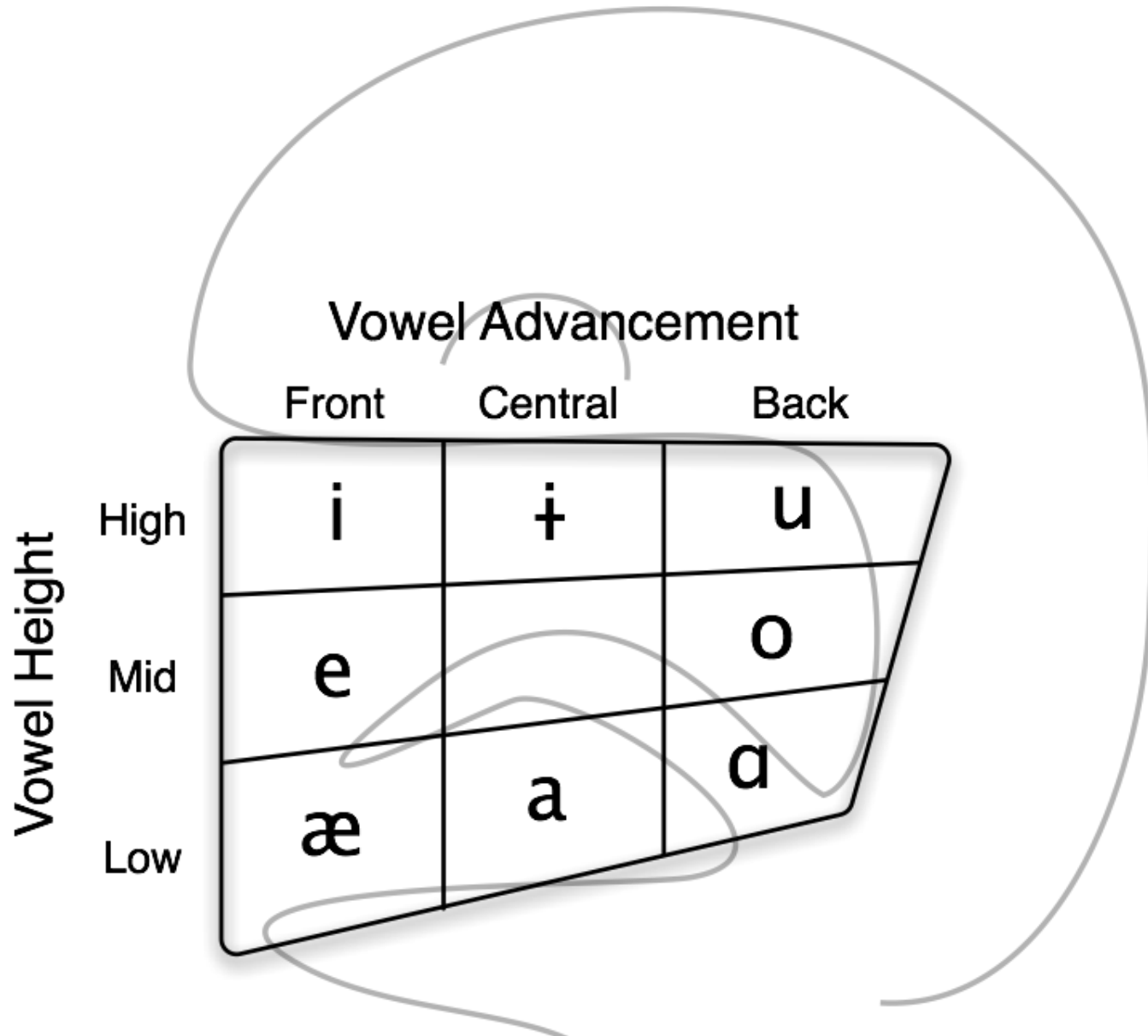
[Assets 2007, International Society of Appraisers \(ISA\) Annual ...](#)  
The International Society of Appraisers (ISA) Annual Convergence on personal property appraising, **Assets** 2005, February 15, 2005 in Chicago IL.  
[www.isa-appraisers.org/conference/index.htm](http://www.isa-appraisers.org/conference/index.htm) - 22k - [Cached](#) - [Similar pages](#) - [Note this](#)

[Fort Worth, TX Video, Site of ASSETS 2007 - Conference of ...](#)  
A Video describing the location of the ASSETS 2007 Conference. Note: turn on your sound.  
Video courtesy of the Fort Worth Convention & Visitors Bureau ...  
[www.isa-appraisers.org/conference/fortworthvideo.htm](http://www.isa-appraisers.org/conference/fortworthvideo.htm) - 18k - [Cached](#) - [Similar pages](#) - [Note this](#)

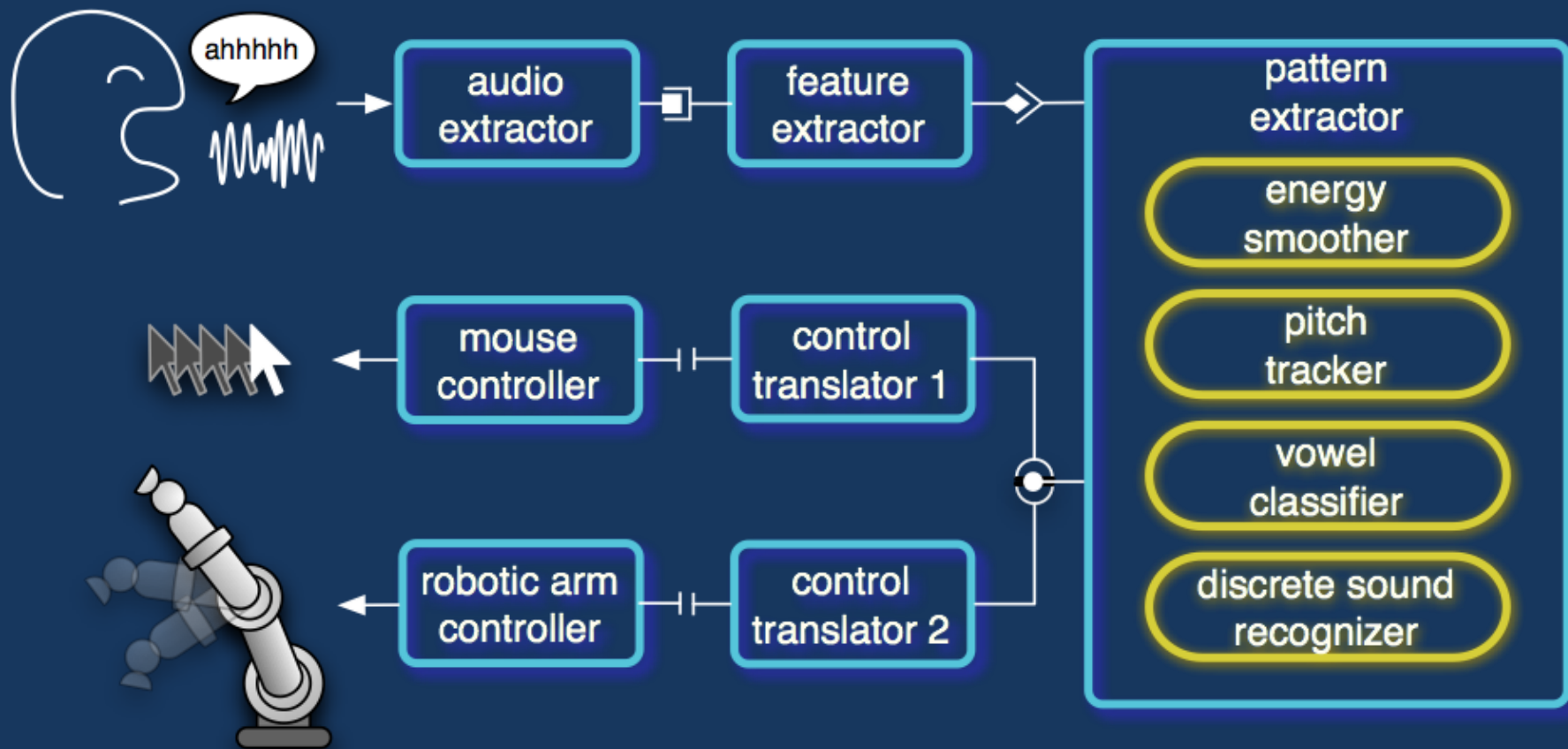
# Vocal Joystick



# Vowel mapping

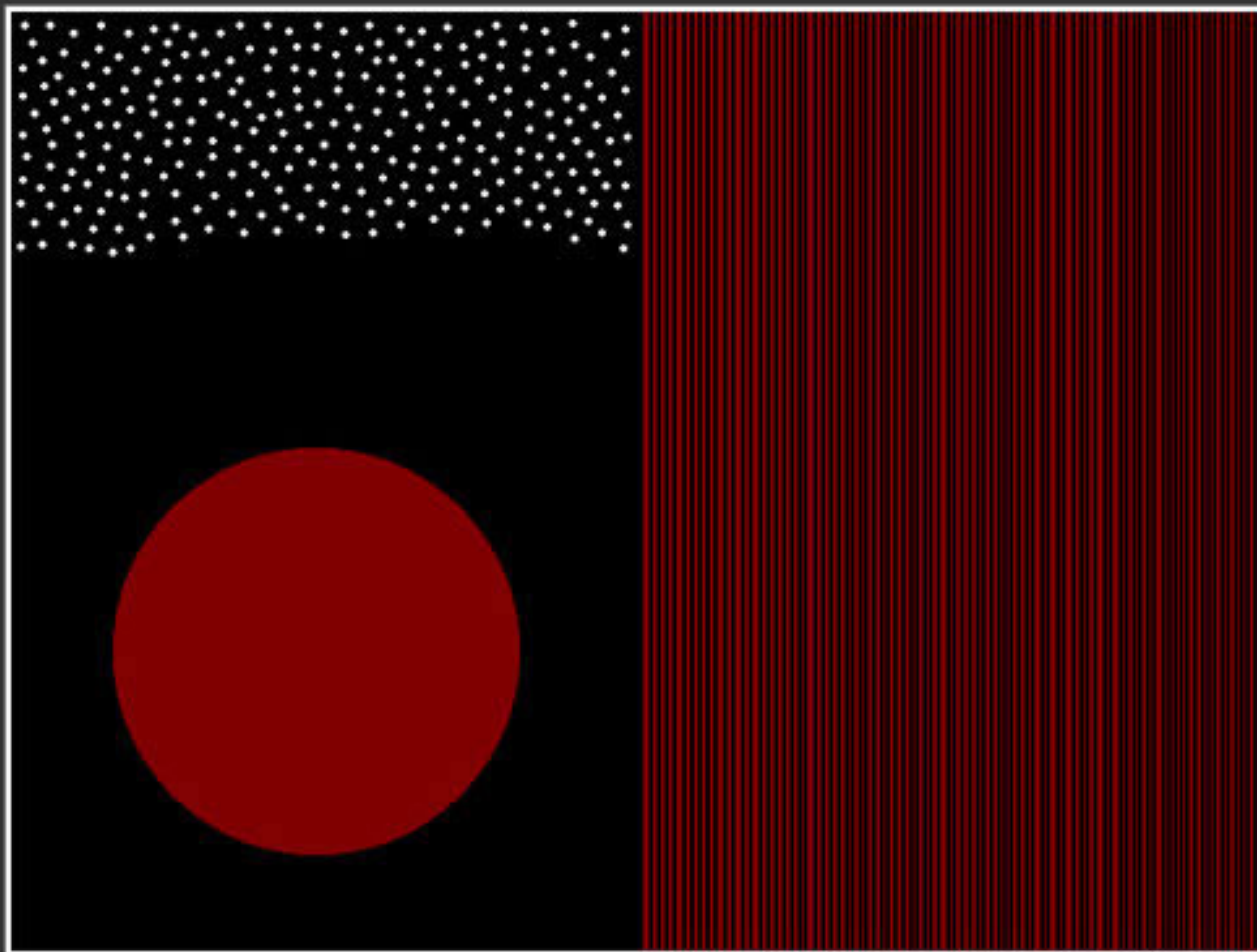


# Vocal Joystick



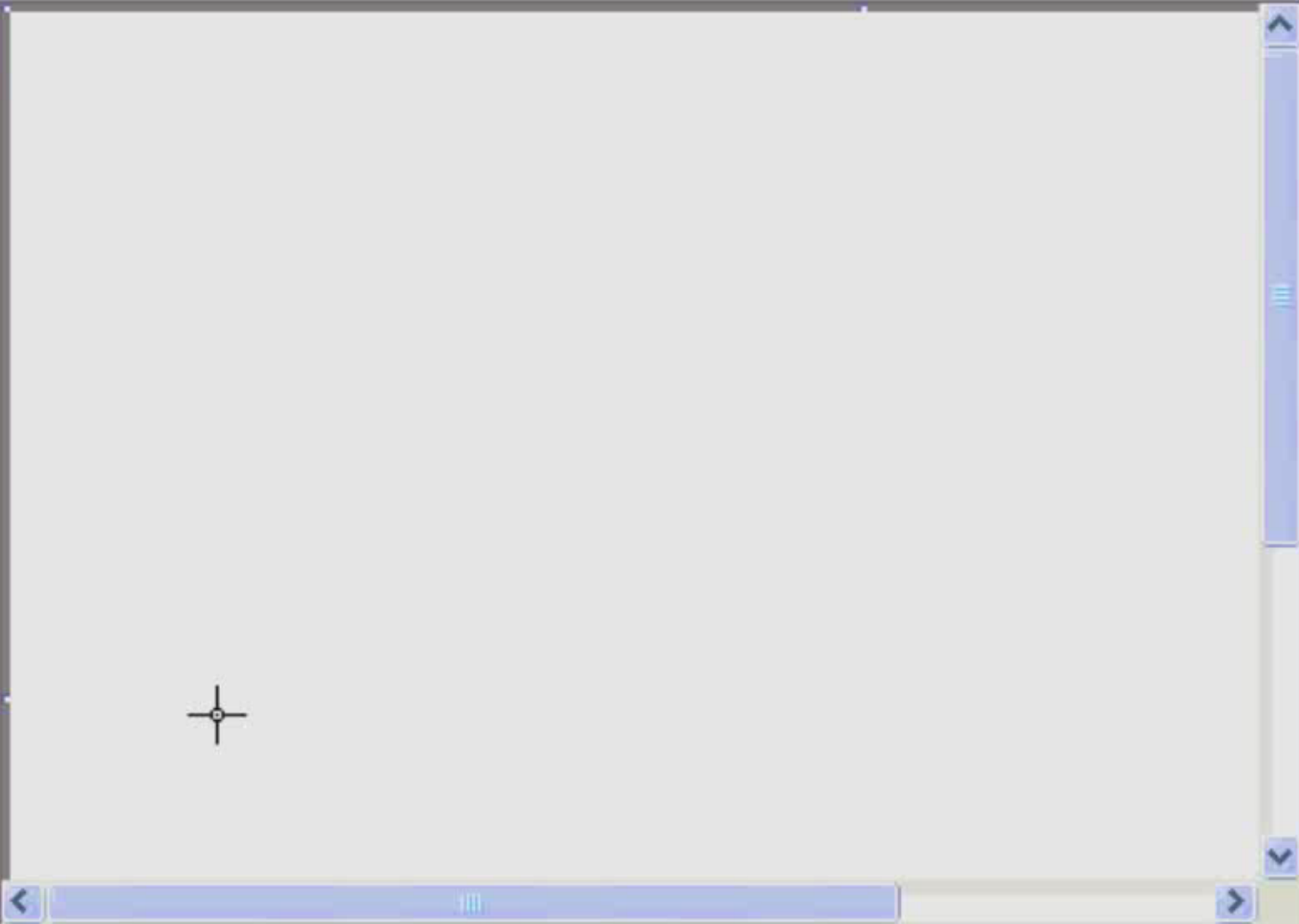


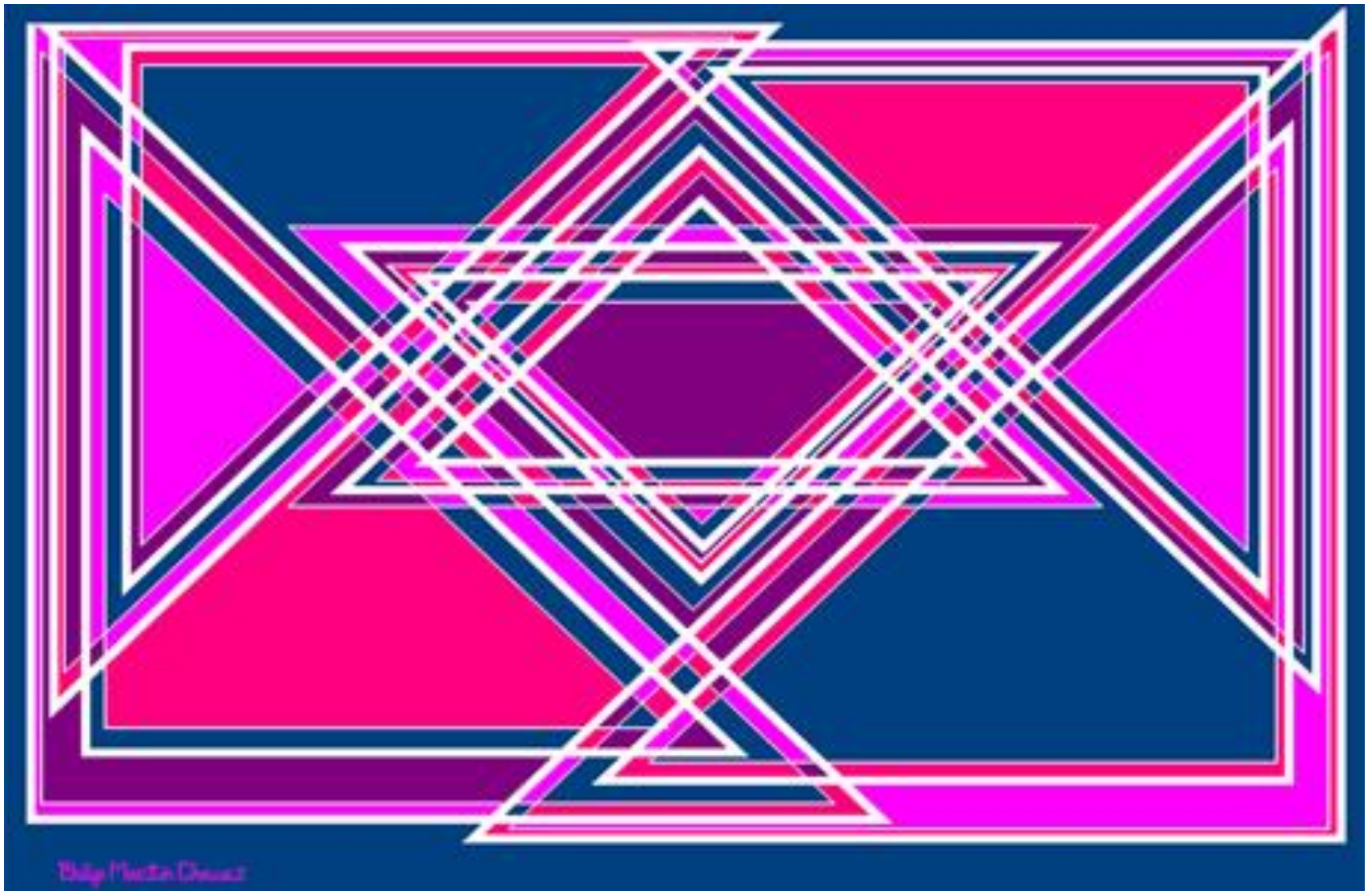
Philip Martin Chavez



Electronic Voice Painter







Big Manta Canvas



Big Manta Canvas





Brush indicator

Loudness meter

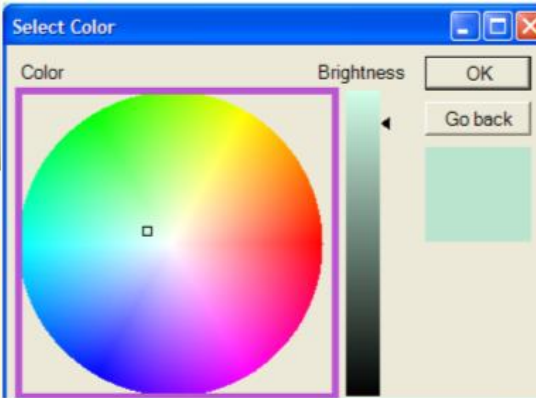
Color  
picker

Help  
overlay

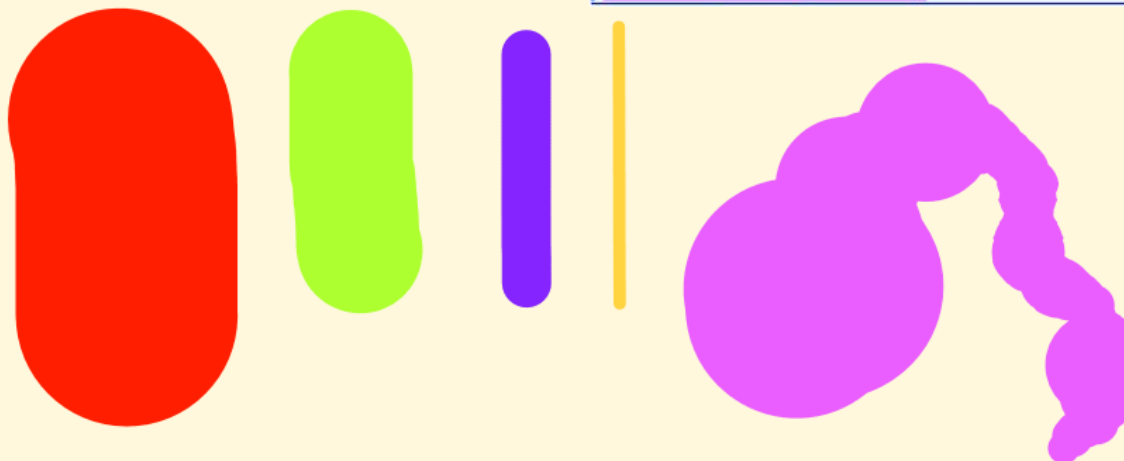
Variable

Speech 

Current mode: MainMenu



Canvas area



"Pie menu"  
"Options"

"Draw"  
"(Very) Slow/Fast"  
"(Very) Thin/Thick"  
"Fixed/Variable thickness"  
"Brush/Background color"  
"Erase"  
"Undo/Redo stroke"

"Clear canvas"  
"Save/Load image"

"Hide/Show help"  
"Go to sleep"  
"Quit program"

Variable



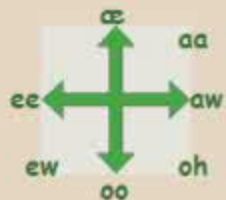
Speech



Current mode: BrushUp

"ck" (Main menu)

"ch" (Brush down)





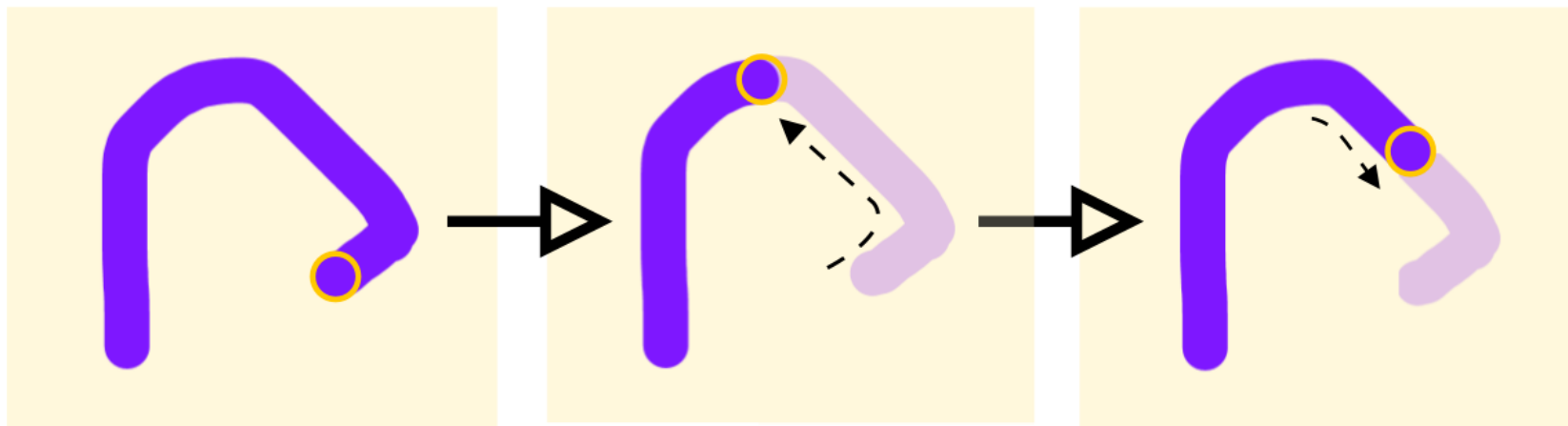
# Continuous undo

"erase"

"aaaa"

"ooo"

"ch"





Speech

Current mode: MainMenu

- "Pie menu"
- "Options"
  
- "Draw"
- "(Very) Slow/Fast"
- "(Very) Thin/Thick"
- "Fixed/Variable thickness"
- "Brush/Background color"
- "Erase"
- "Undo/Redo stroke"
  
- "Clear canvas"
- "Save/Load image"

---

- "Hide/Show help"
- "Go to sleep"
- "Quit program"

Fixed



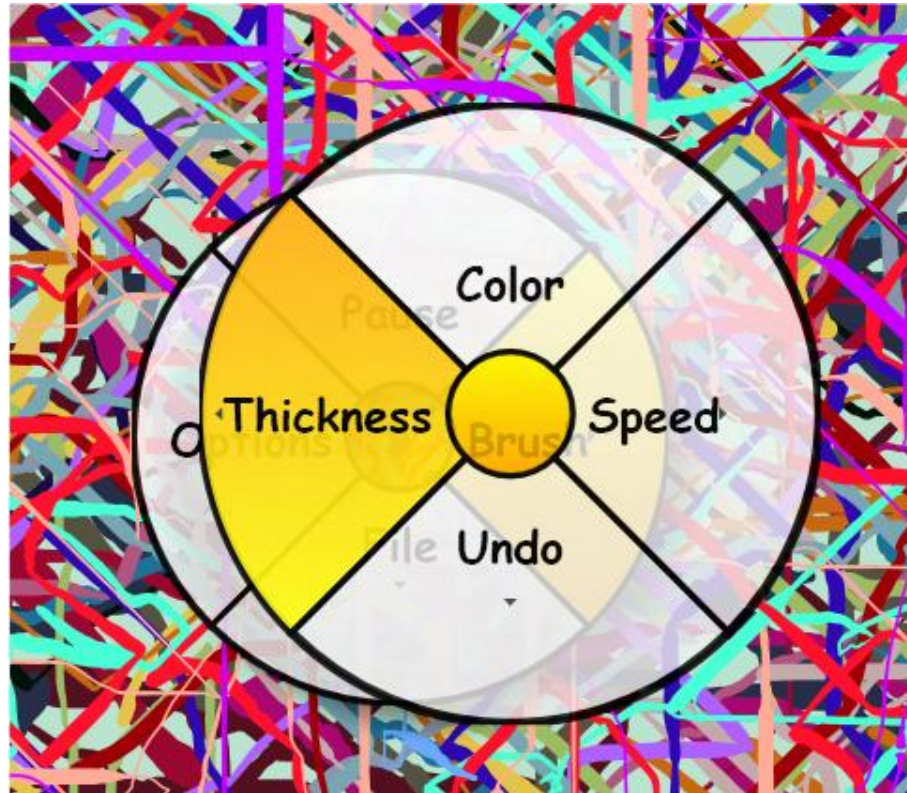
Speech

Current mode: Start

"Wake up"

# Voice marking menu

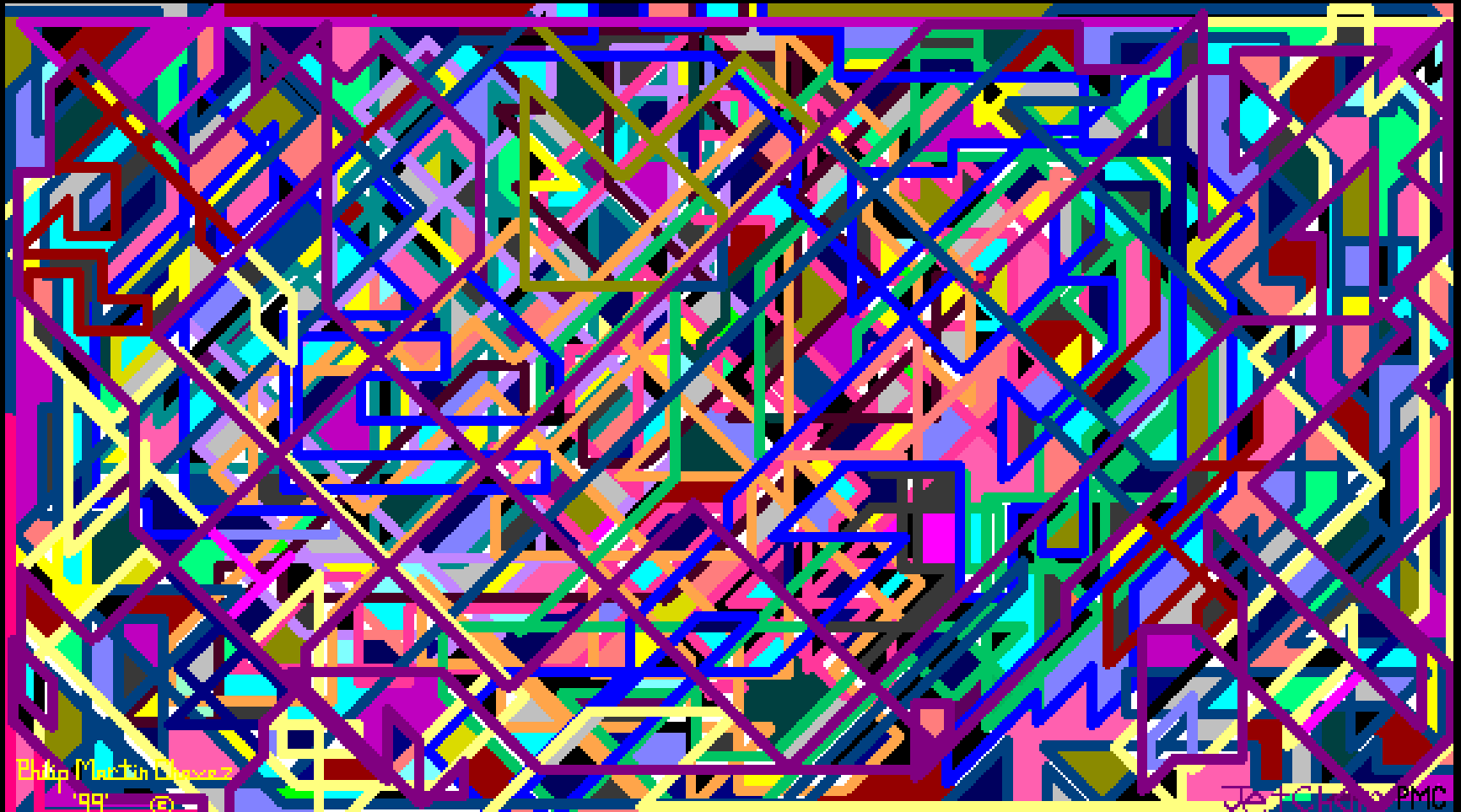
“ch” “d”



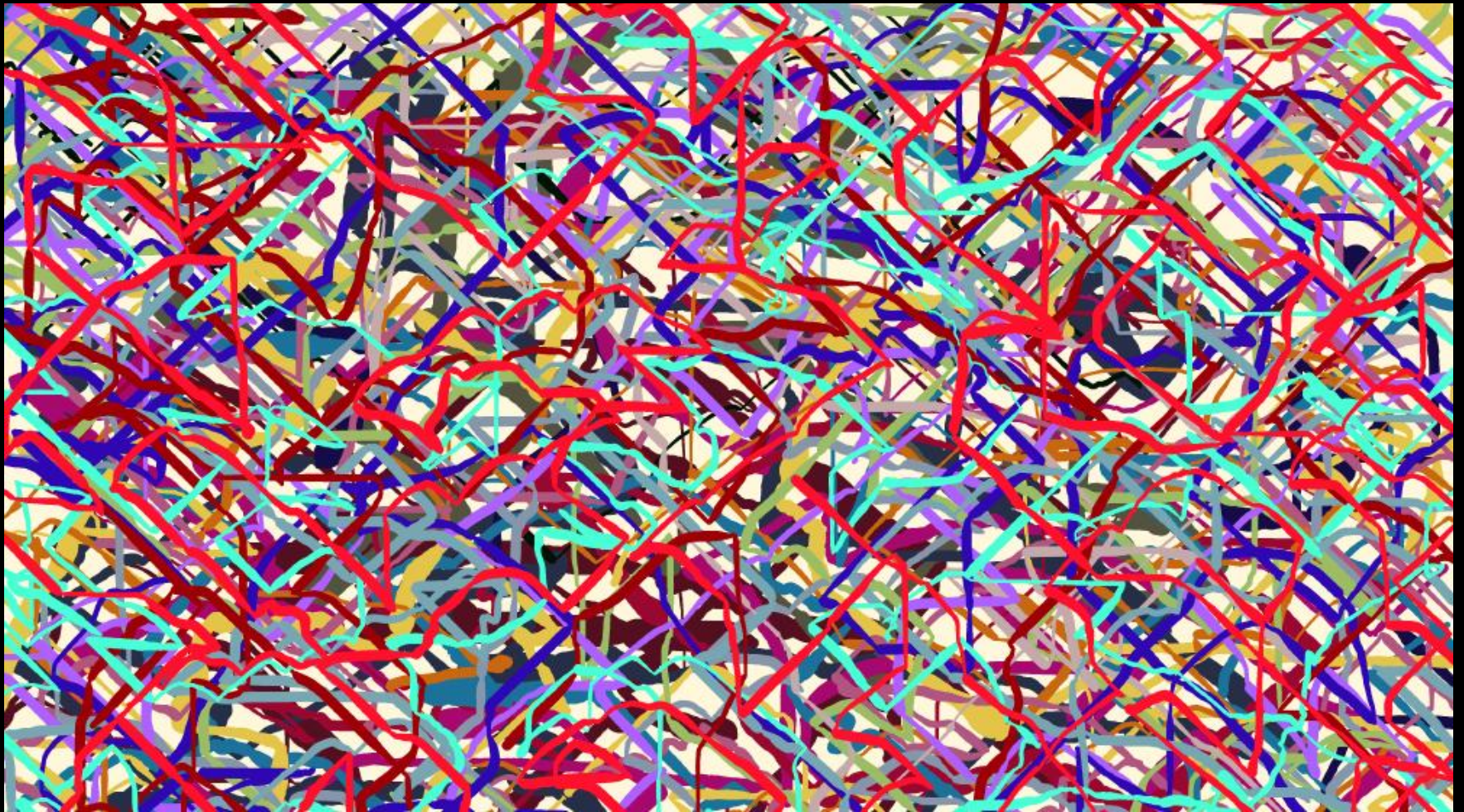
Created using VoiceDraw (2.5 hours)



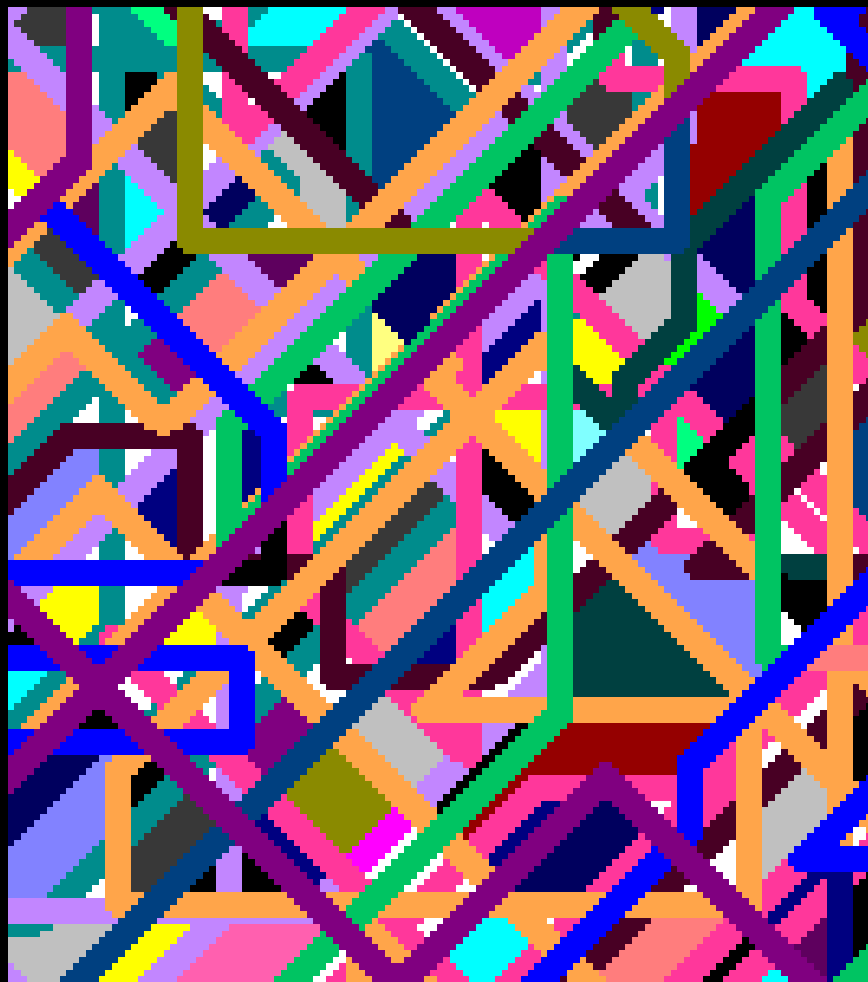
Created using MS Paint & Dragon Dictate (9 hours)



Created using VoiceDraw (3 hours)



Dragon Dictate & MS Paint



VoiceDraw





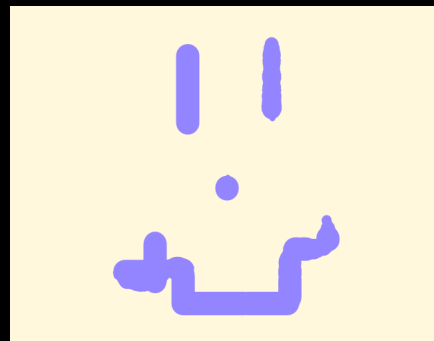
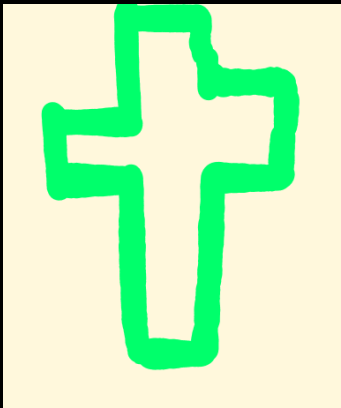
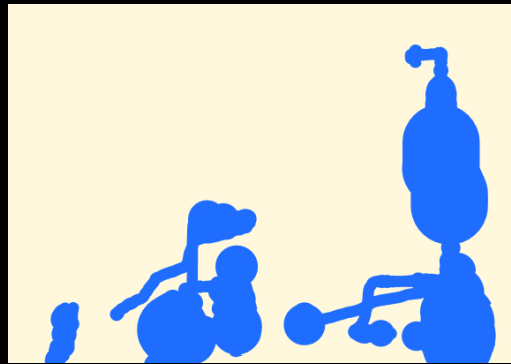
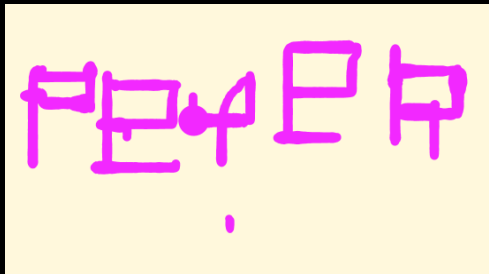
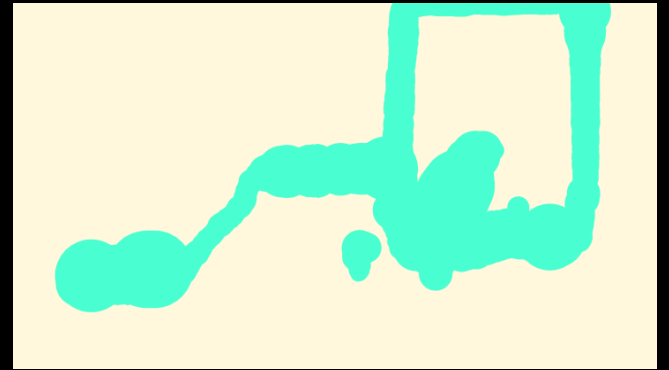
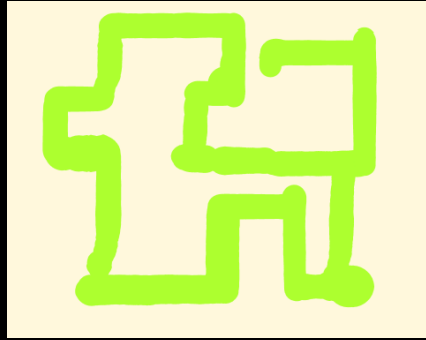


**The Vocal Joyflick: VoiceBot**  
The Vocal Joyflick: VoiceBot is a small robot that can be programmed to speak and perform simple tasks. It is a great tool for learning about robotics and programming.

**How does it work?**  
The robot is controlled by a computer program. The program sends signals to the robot's motor, which makes it move. The robot also has a microphone and a speaker, so it can hear and speak.

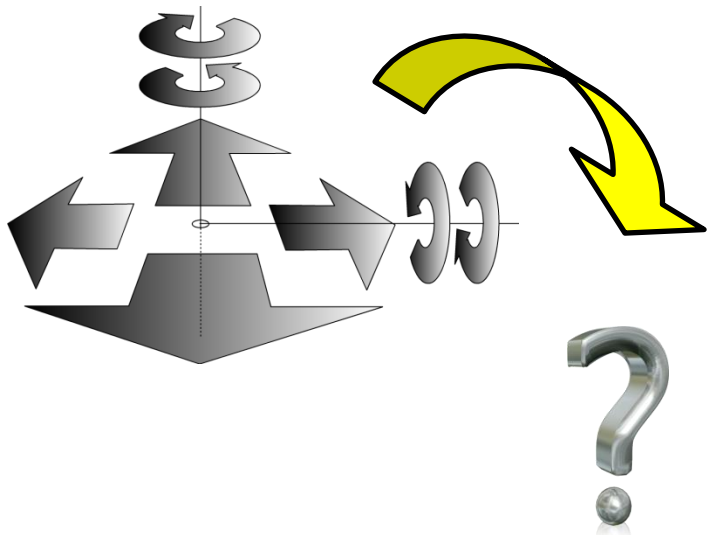
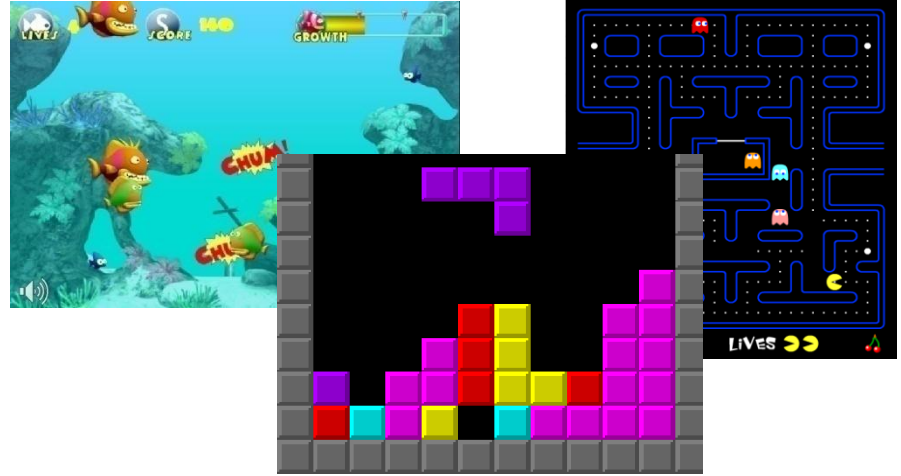
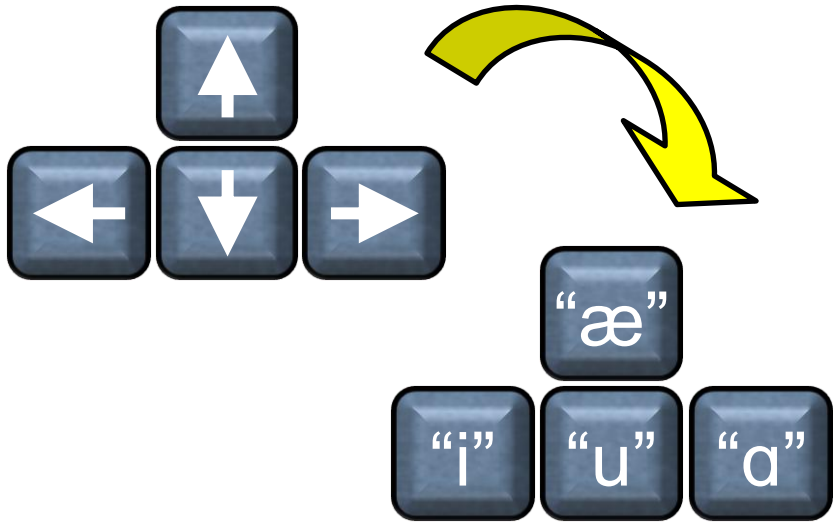
**What can it do?**  
The robot can be programmed to perform a variety of tasks, such as moving in a straight line, turning, and speaking. It can also be programmed to respond to voice commands.

**Why is it called "The Vocal Joyflick: VoiceBot"?**  
The robot is called "The Vocal Joyflick: VoiceBot" because it can speak and perform simple tasks. It is a great tool for learning about robotics and programming.



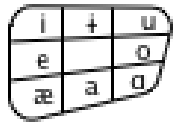


# VoiceGames



# Vocal parameters

## vocal input parameters



vowels

## abstract value types

(a)

continuous cyclic 1-D



(b)

discrete 0-D



1 point

discrete 1-D



(d)

multiple points

(e)

continuous linear 1-D



(f)

## control parameter types

(g)

angular position



(h)

Δ position in 2-D



(i)

switch toggle



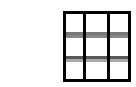
(j)

discrete Δ in 1-D



(k)

discrete value selection



(l)

scroll 1-D



(m)

variable control



loudness



pitch

# Thank you! Questions! Discussions!

Susumu Harada

harada@cs.washington.edu

<http://www.vocaljoystick.org>

