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

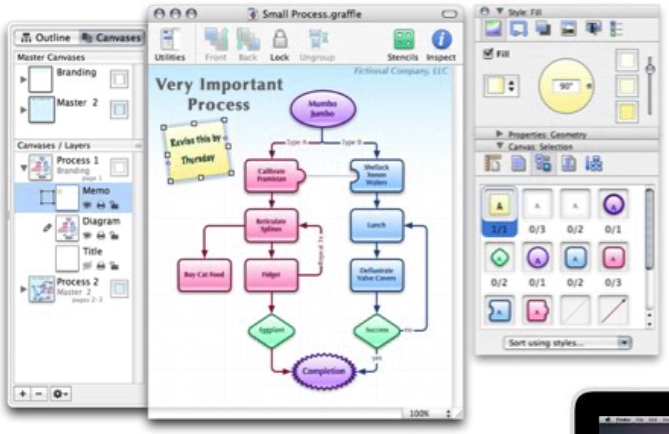
Xiao Li
Electrical Engineering (now at MSR)

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

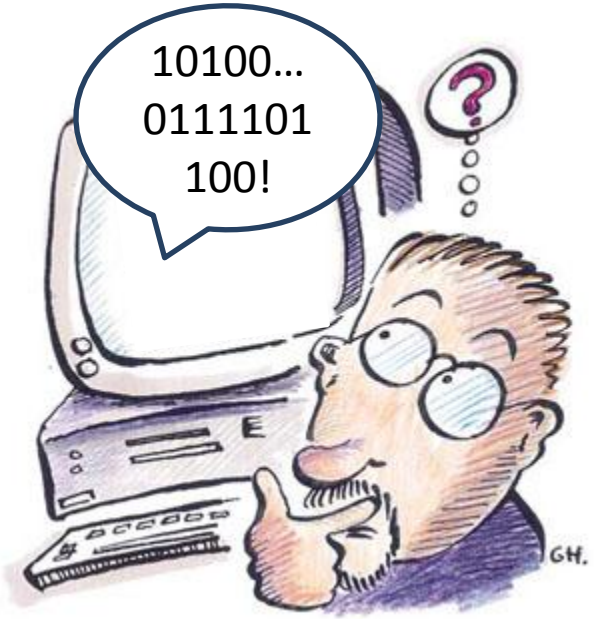
250,000	People with spinal cord injury
47%	are quadriplegic
70%	are unemployed
46,000,000	Adults diagnosed with arthritis
1,000,000	Adults with Parkinson's disease
50,000	Children and adults with Muscular dystrophy

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

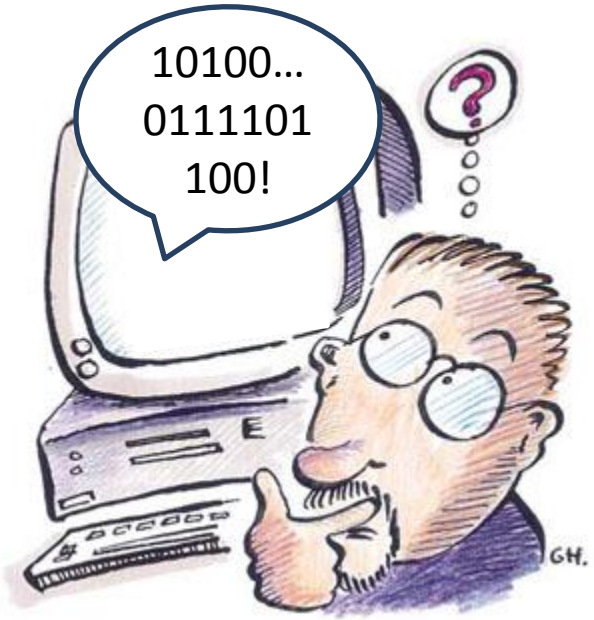
http://www.microsoft.com/enable/aging/scan3_large.aspx



http://www.ala.org/ala/alonline/resources/selectedarticles/10reasonswhy.cfm

Output

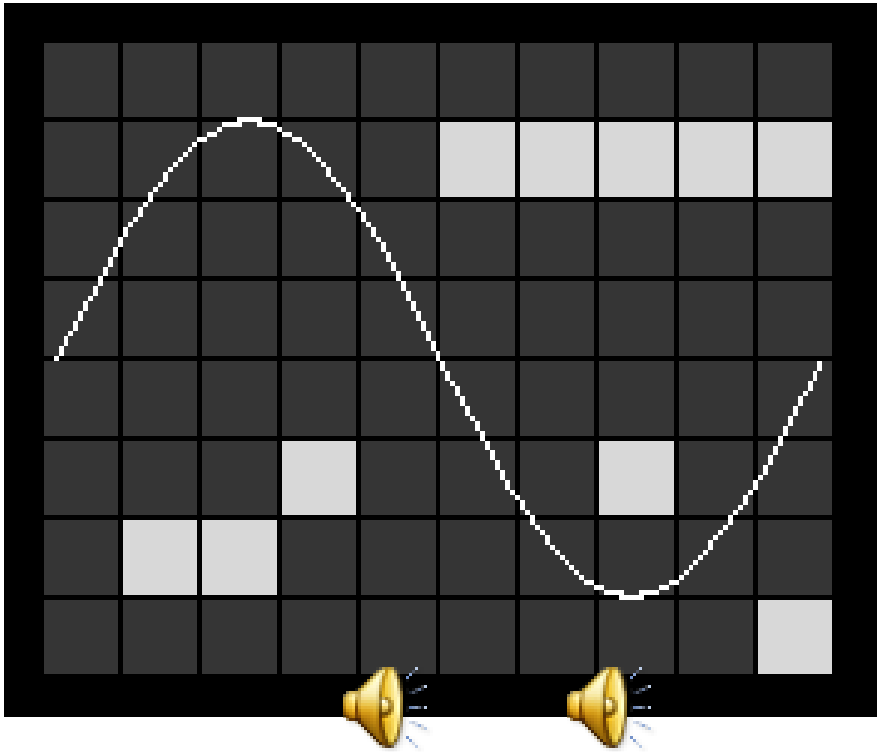
- The vOICe
- PLUMB
- 3D Audio Web Browser
- Games for the Blind



http://www.ala.org/ala/alonline/resources/selectedarticles/10reasonswhy.cfm

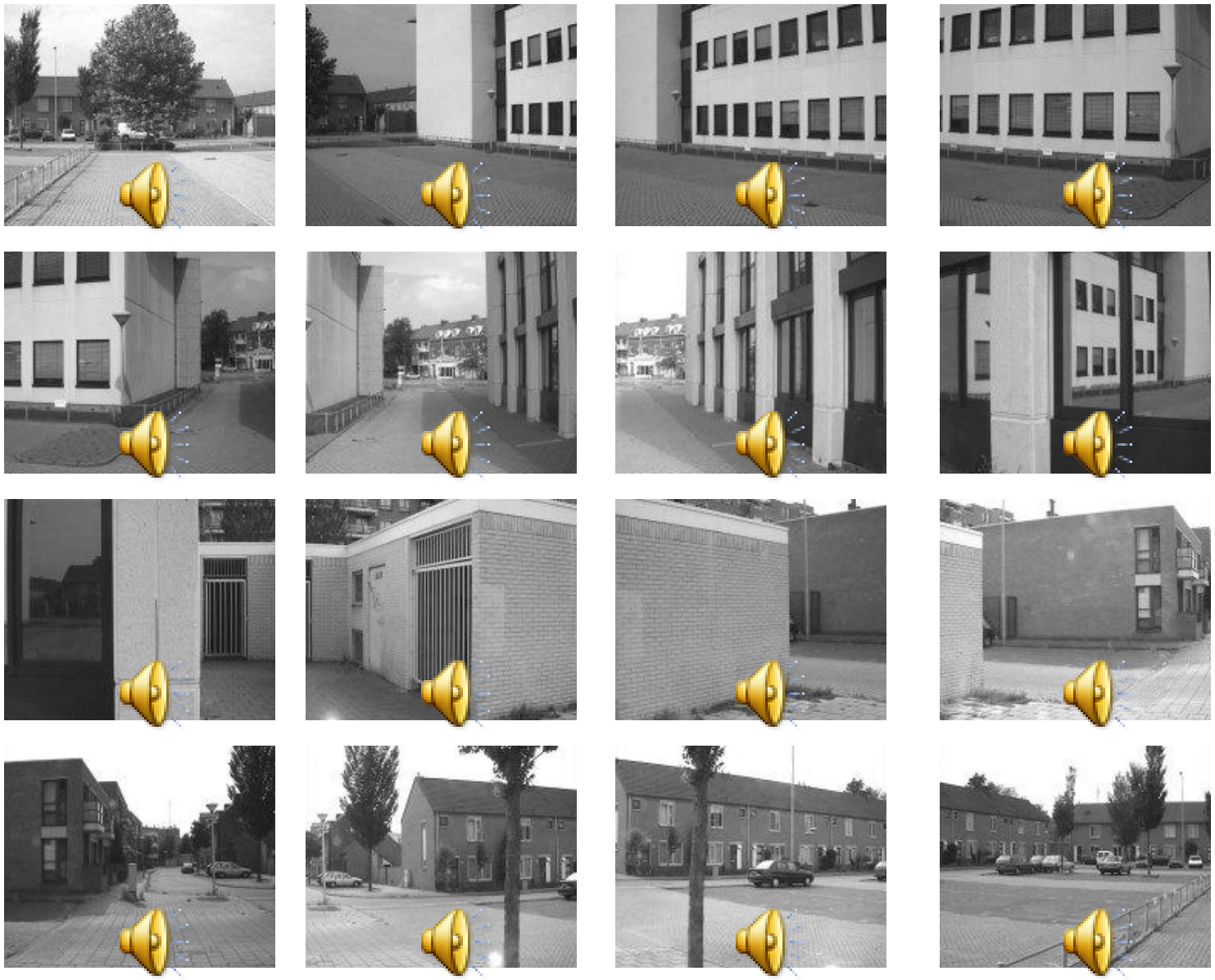
The vOICe

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



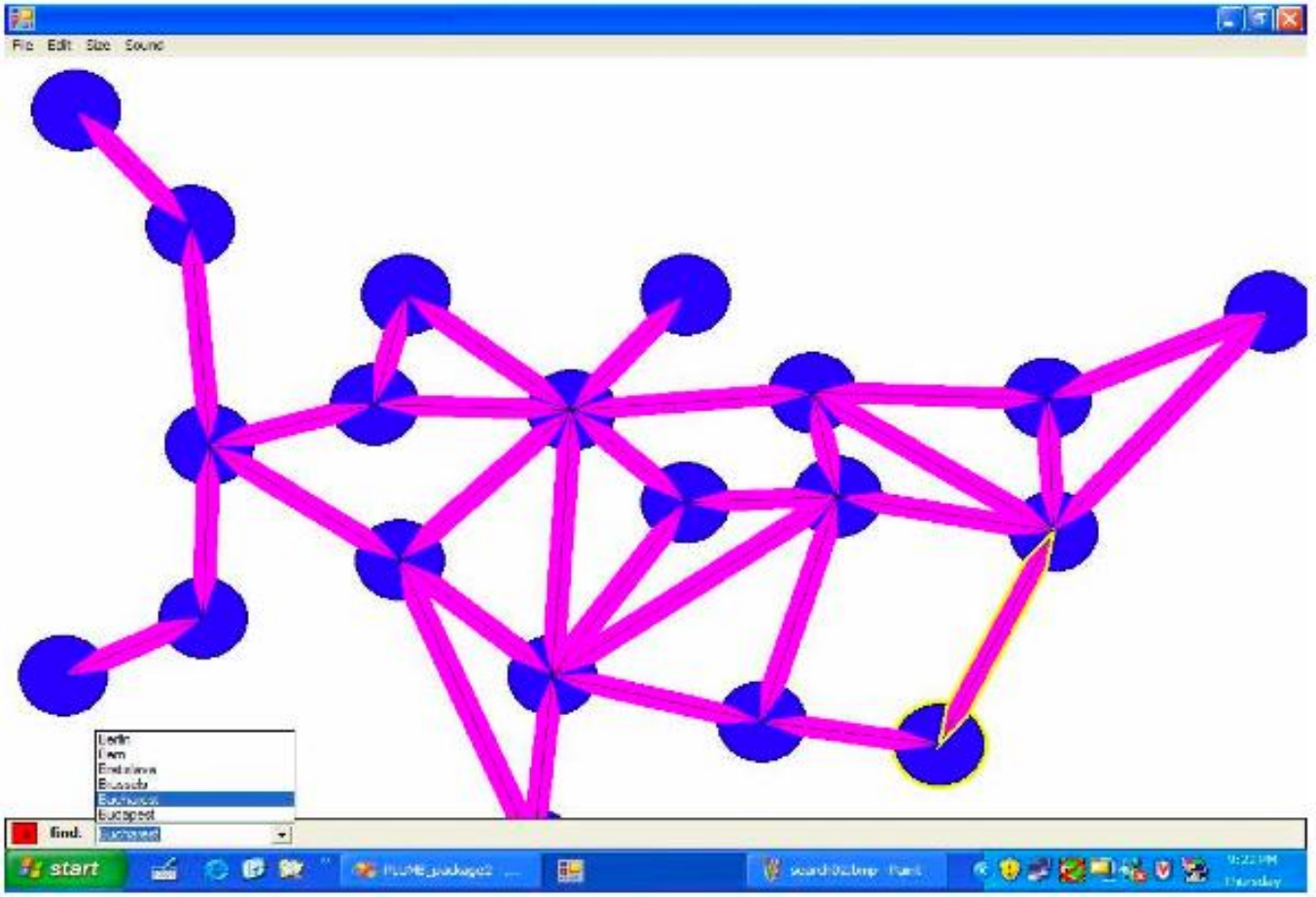
The vOICe

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



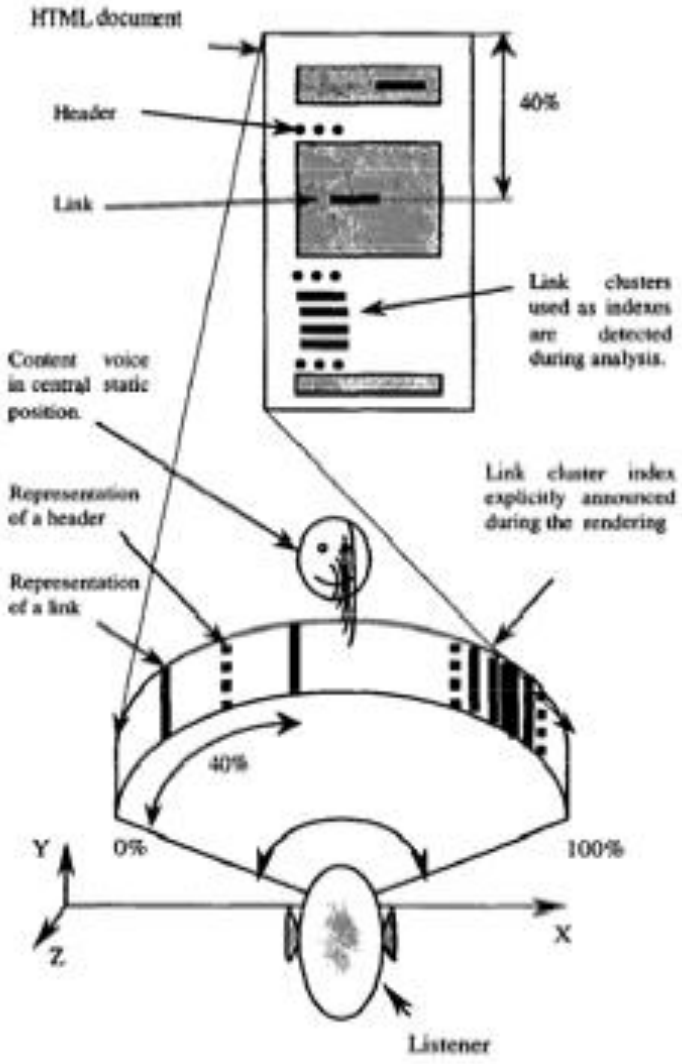
PLUMB

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



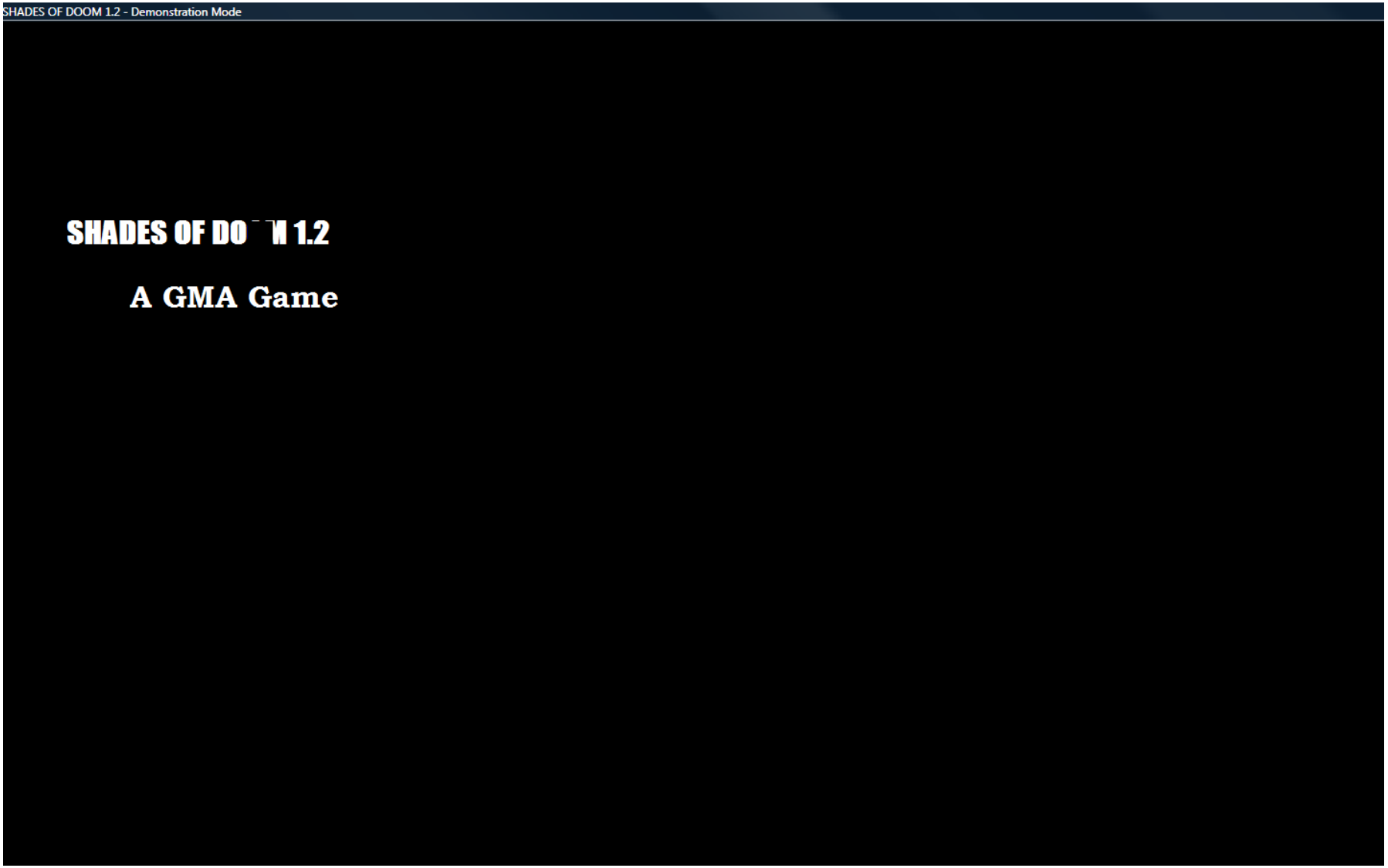
3D audio

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

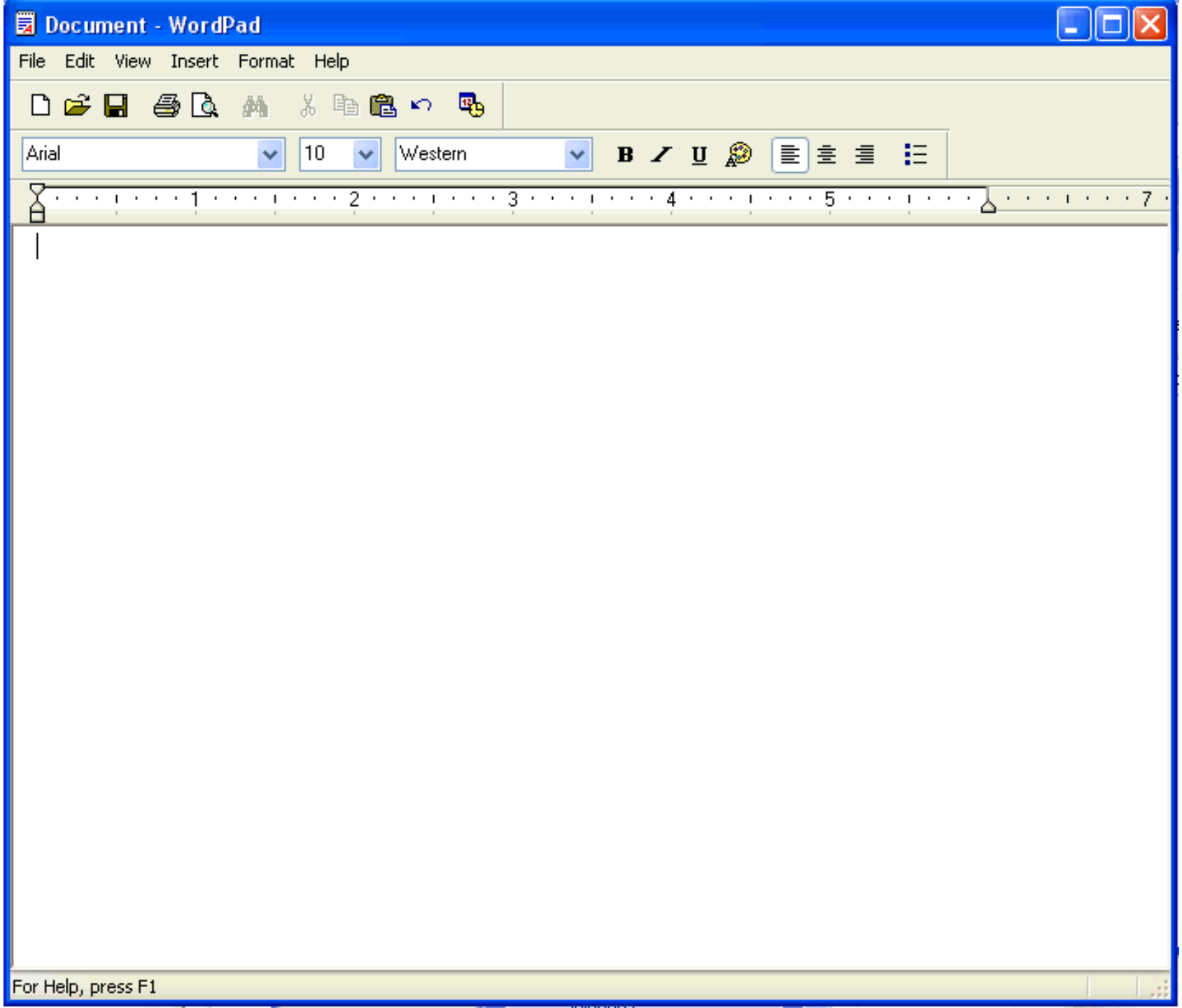


Voice Input

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.


Document - WordPad
File Edit View Insert Format Help
Arial
10
Western
For Help, press F1

Filled pause

Masataka Goto, Katunobu Itou, and Satoru Hayamizu.
A Real-time Filled Pause Detection System for Spontaneous Speech Recognition.
Eurospeech 1999

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 spotter

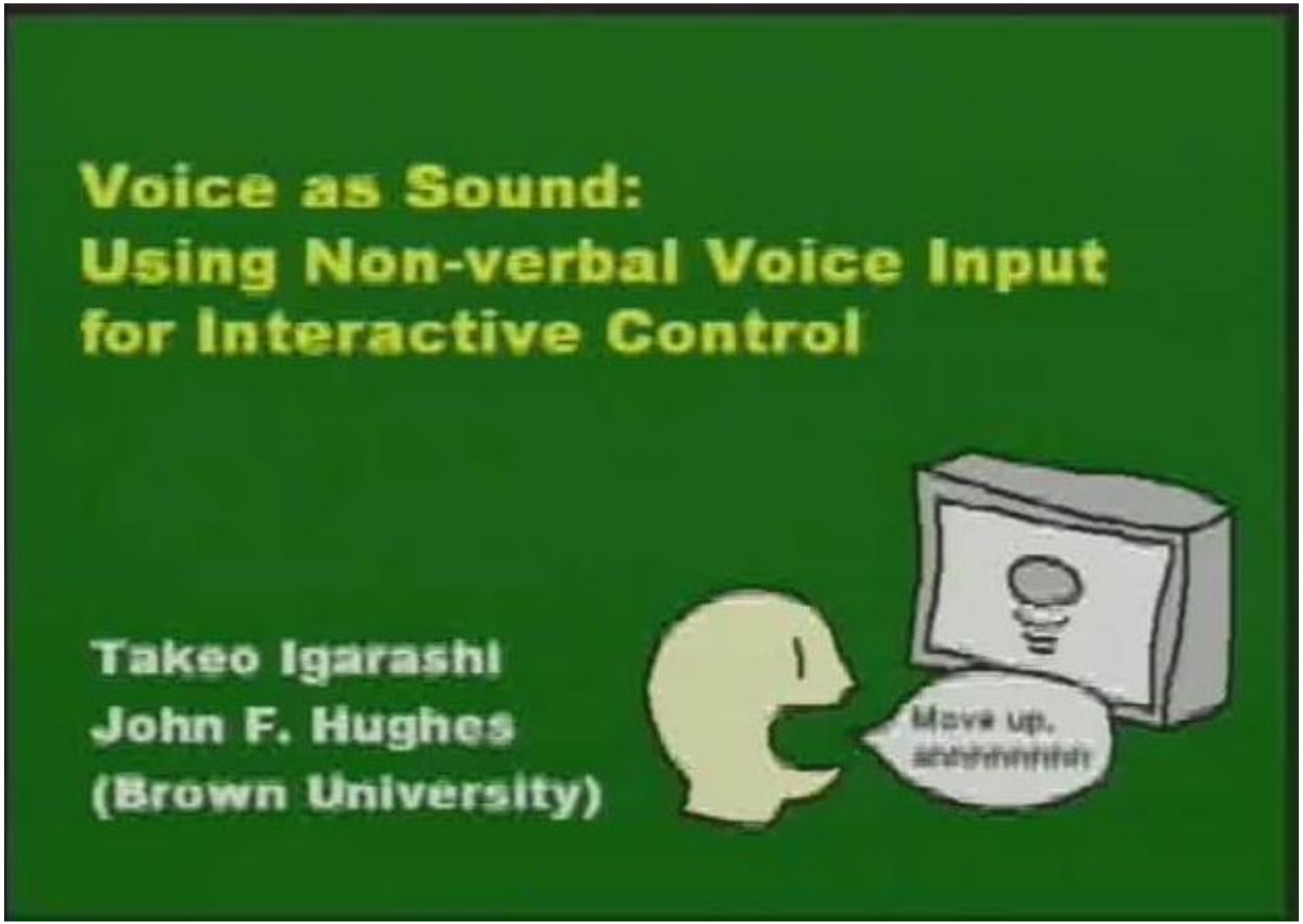
Masataka Goto, Koji Kitayama, Katunobu Itou, and Tetsunori Kobayashi
Speech Spotter: On-demand Speech Recognition in Human-Human Conversation on the Telephone or in Face-to-Face Situations.
ICSLP 2004

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

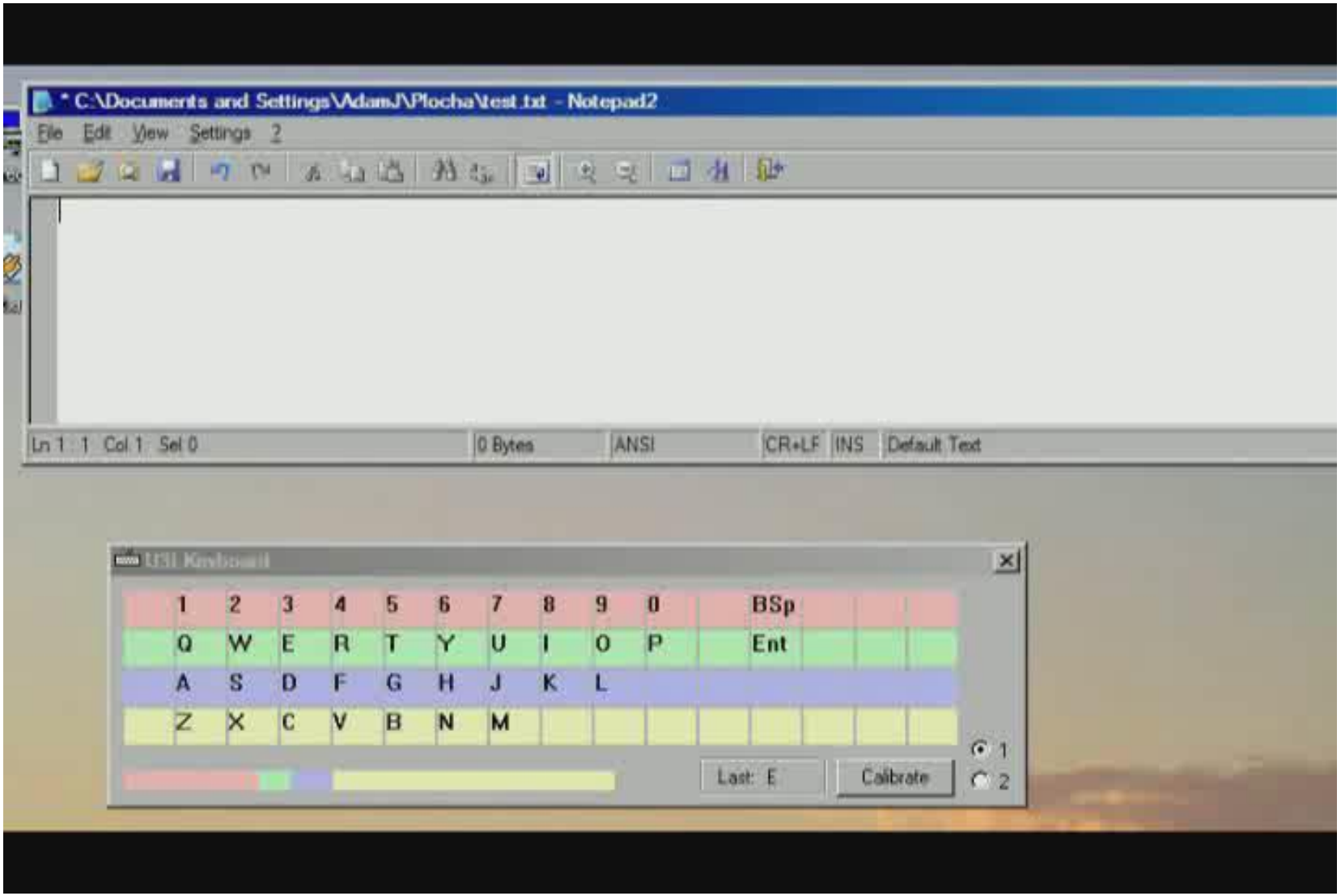
Voice as sound

Takeo Igarashi and John F. Hughes.
Voice as Sound: Using Non-verbal Voice Input for Interactive Control.
UIST 2001



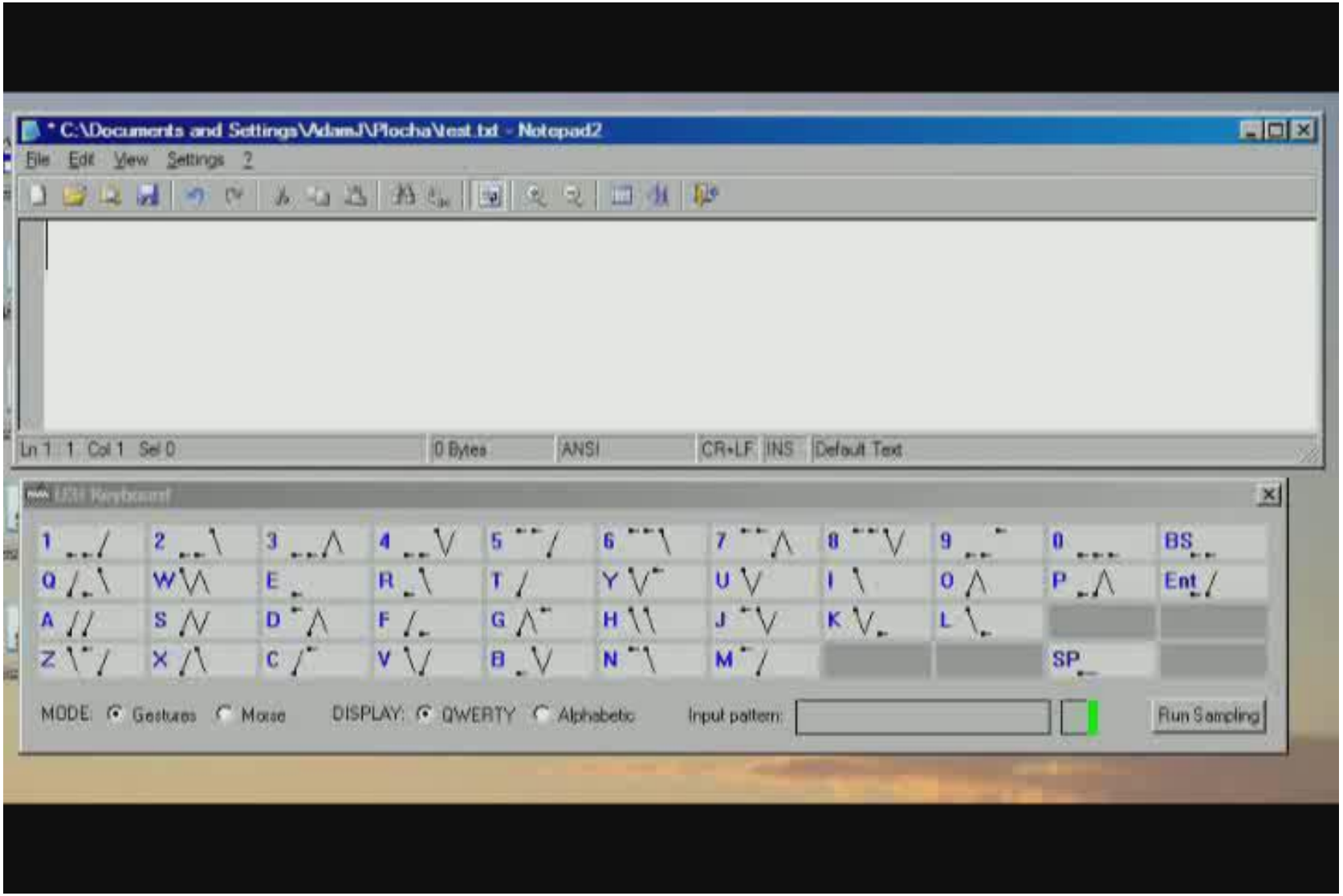
Whistle text entry

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



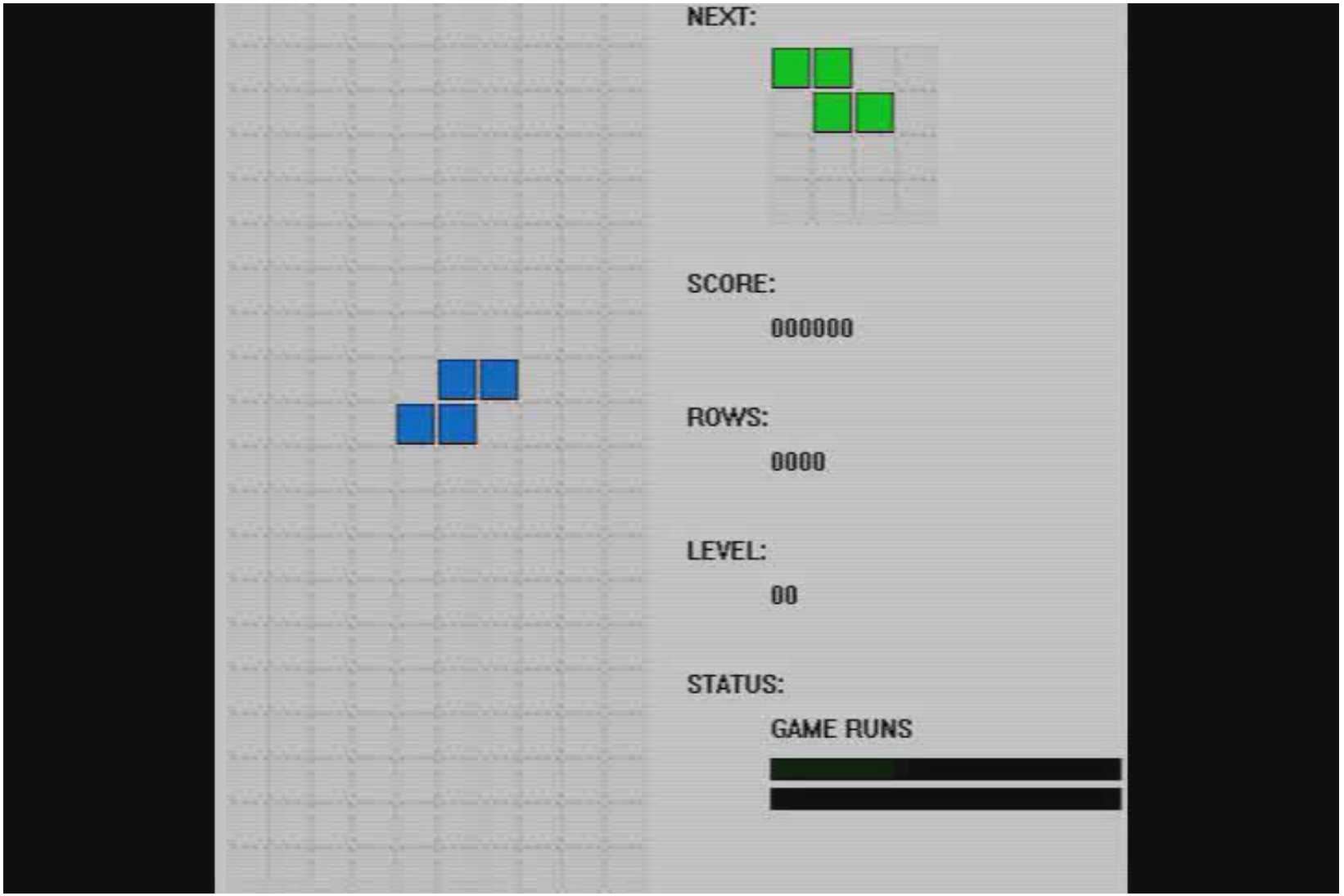
Whistle text entry

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

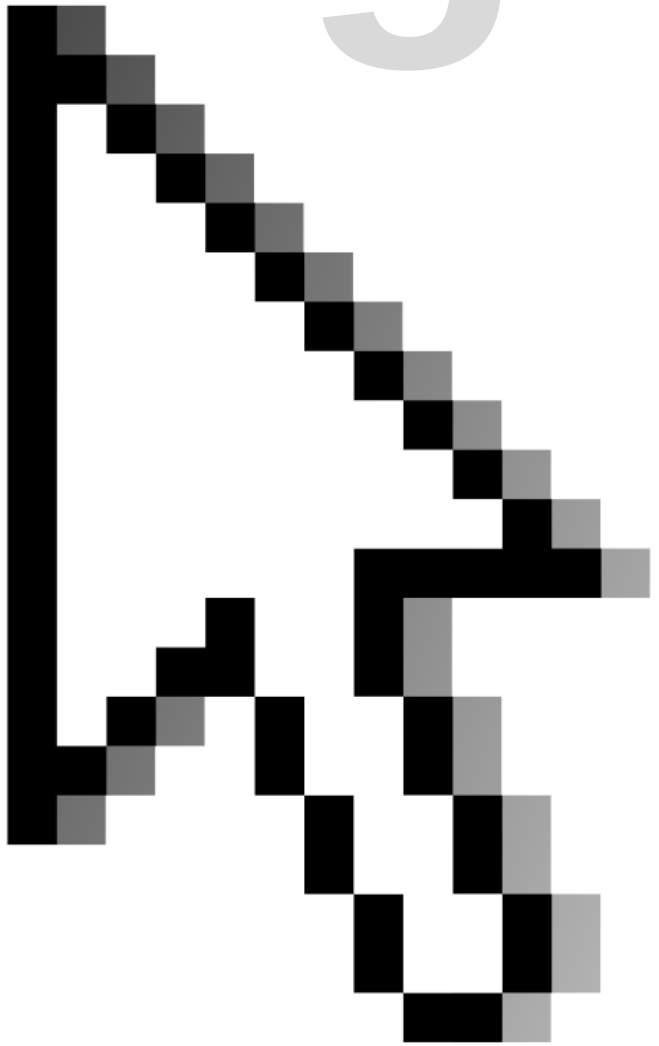


Humming game

Adam Sporka, Sri Kurniawan, Muni Mahmud, and Pavel Slavik.
Non-speech Input vs Speech Recognition: Real-time Control of Computer Games.
ASSETS, 2006



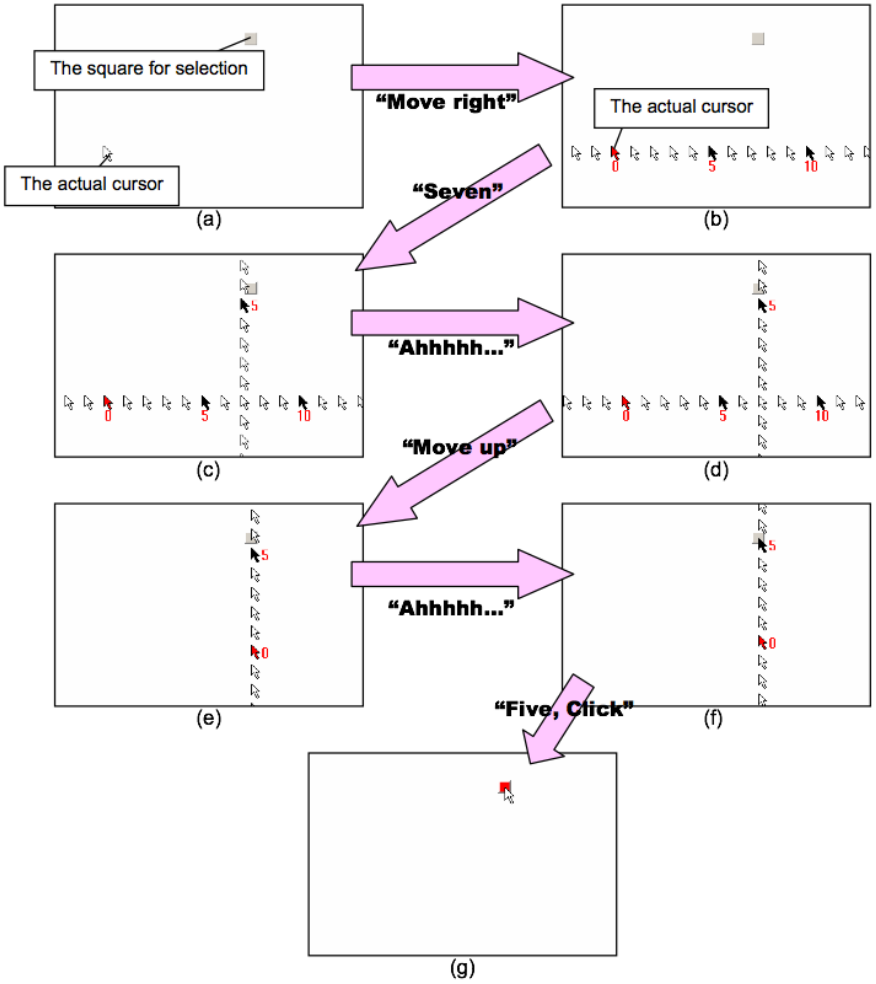
Pointing with Voice



Dragon Dictate

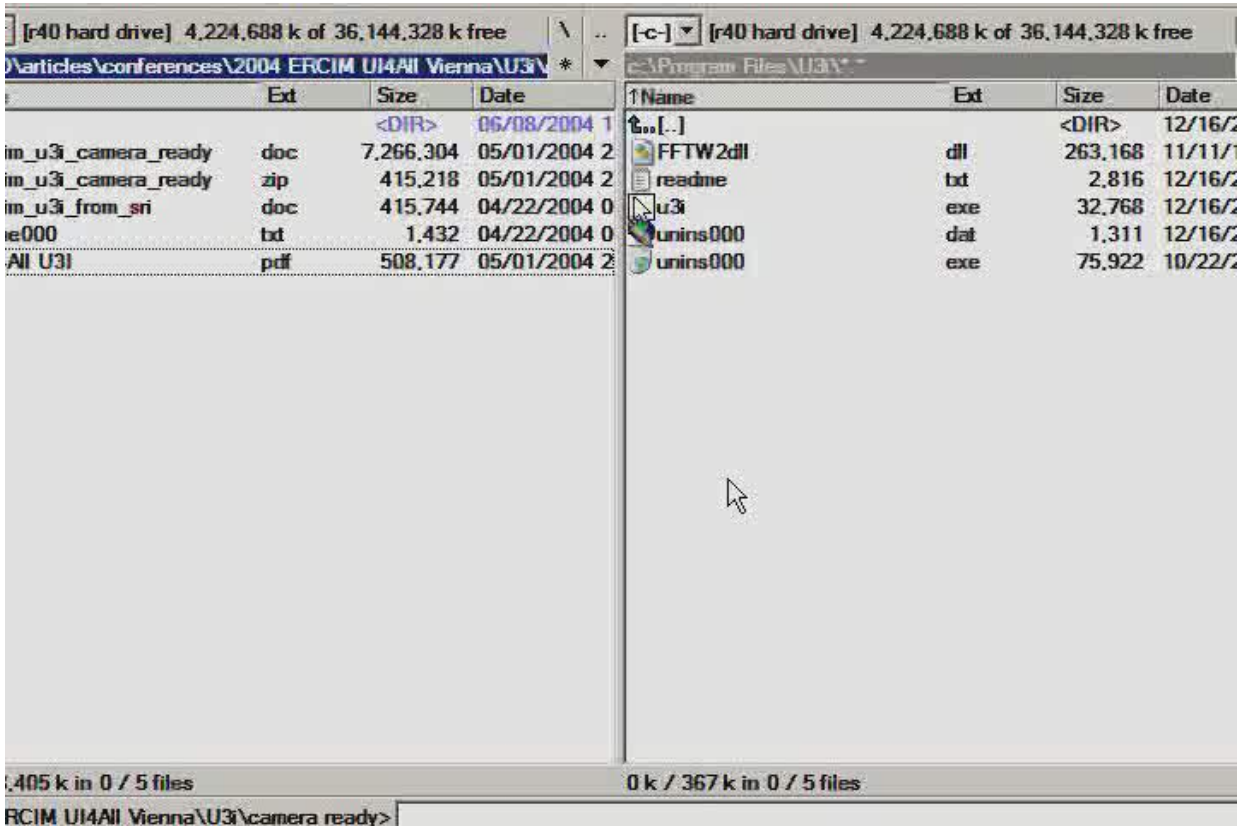
Migratory cursor

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



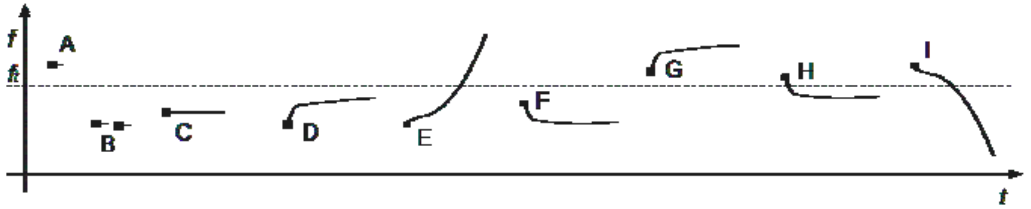
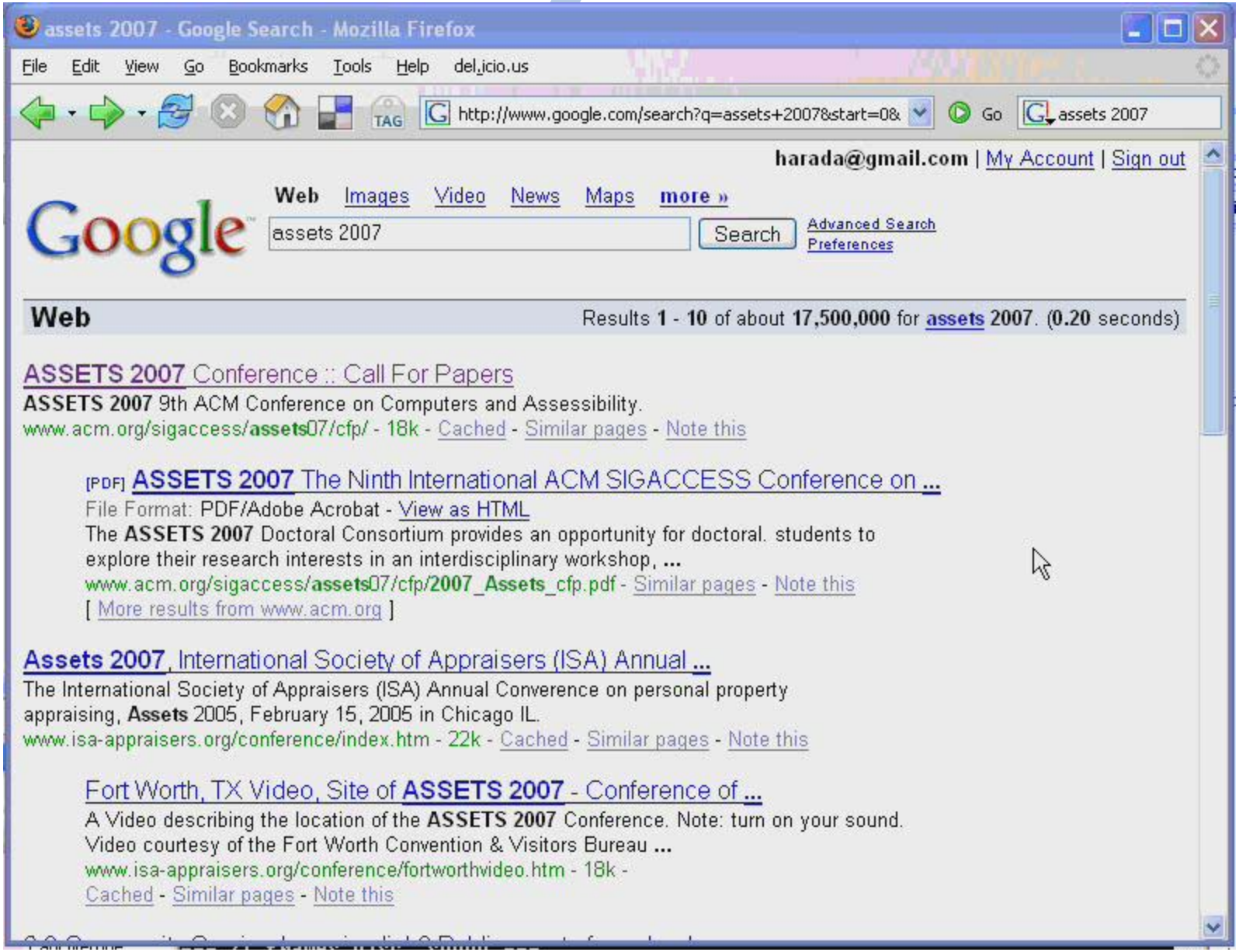
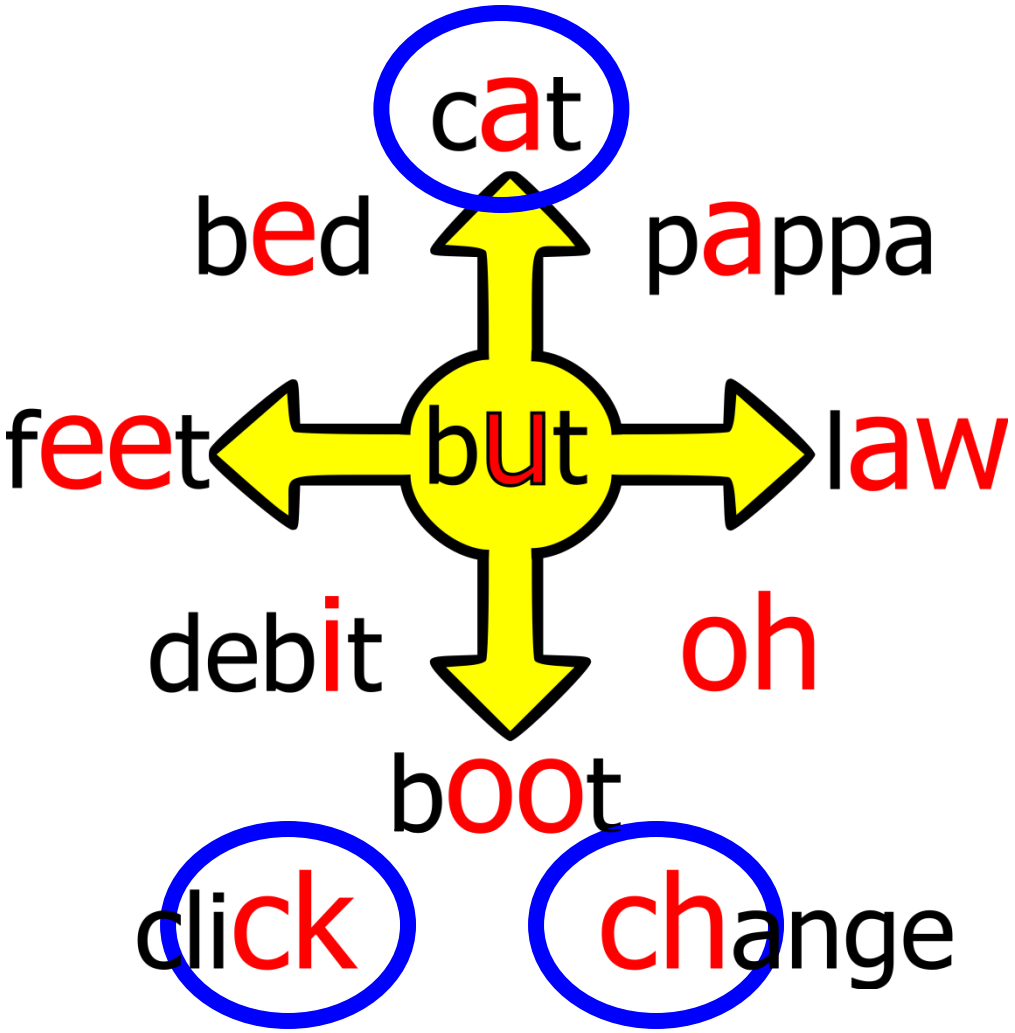


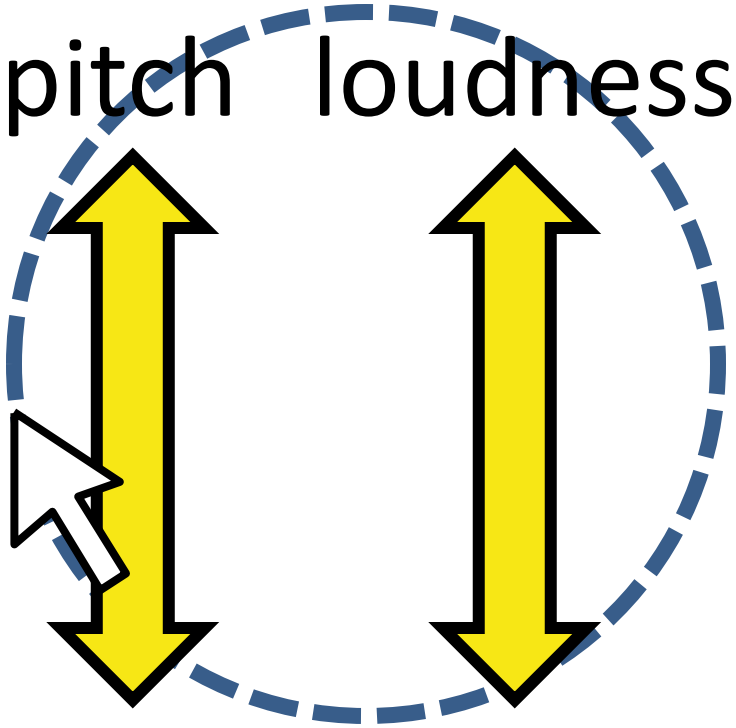
Fig. 2. Examples of control tones. t – time, f – pitch, ft – threshold pitch; A – click, B – double click, C – no motion, D – motion to the right, E – fast motion to the right, F – motion to the left, G – motion up, H – motion down, I – fast motion down.

Vocal Joystick

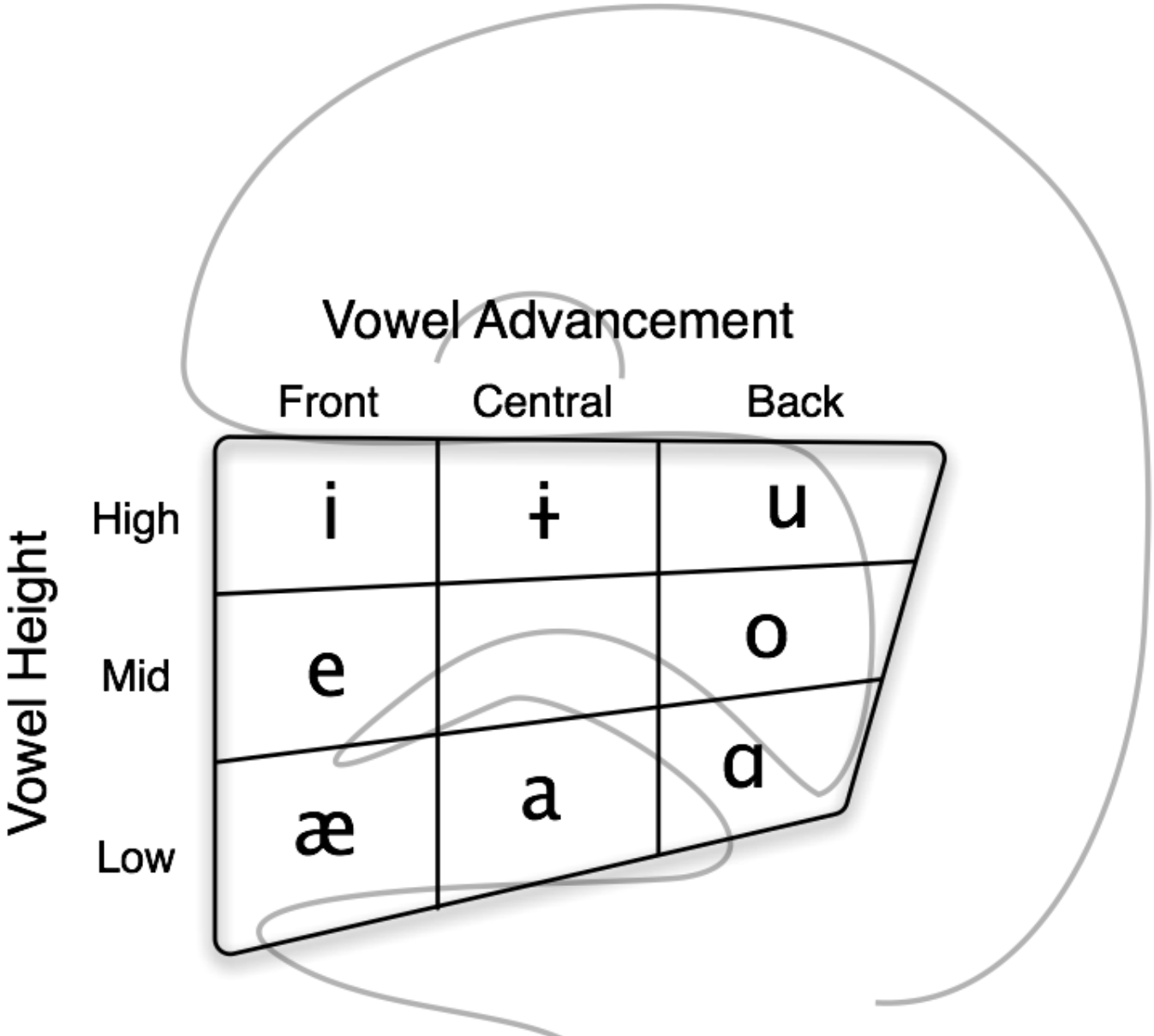


Vocal Joystick

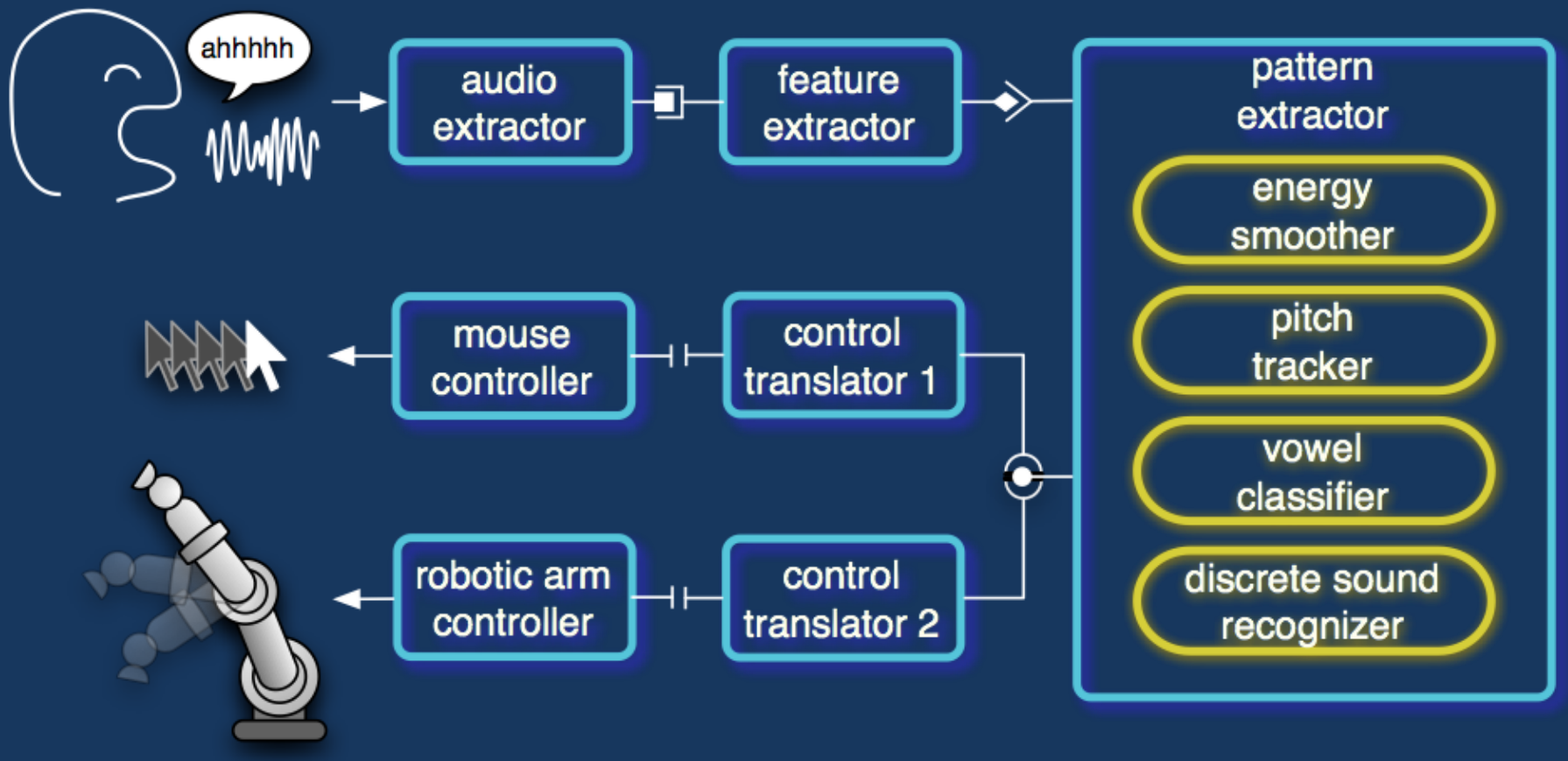




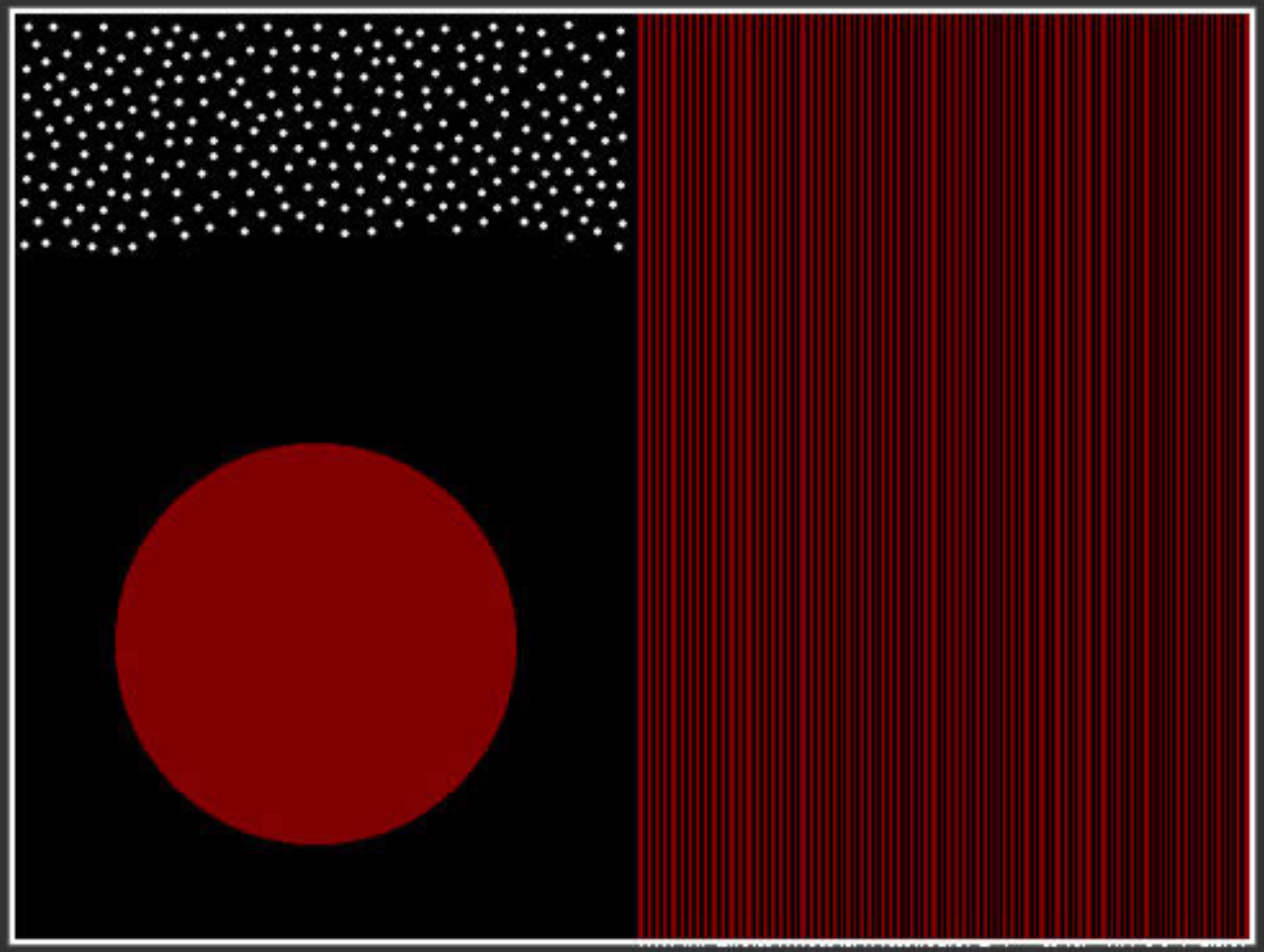
Vowel mapping



Vocal Joystick

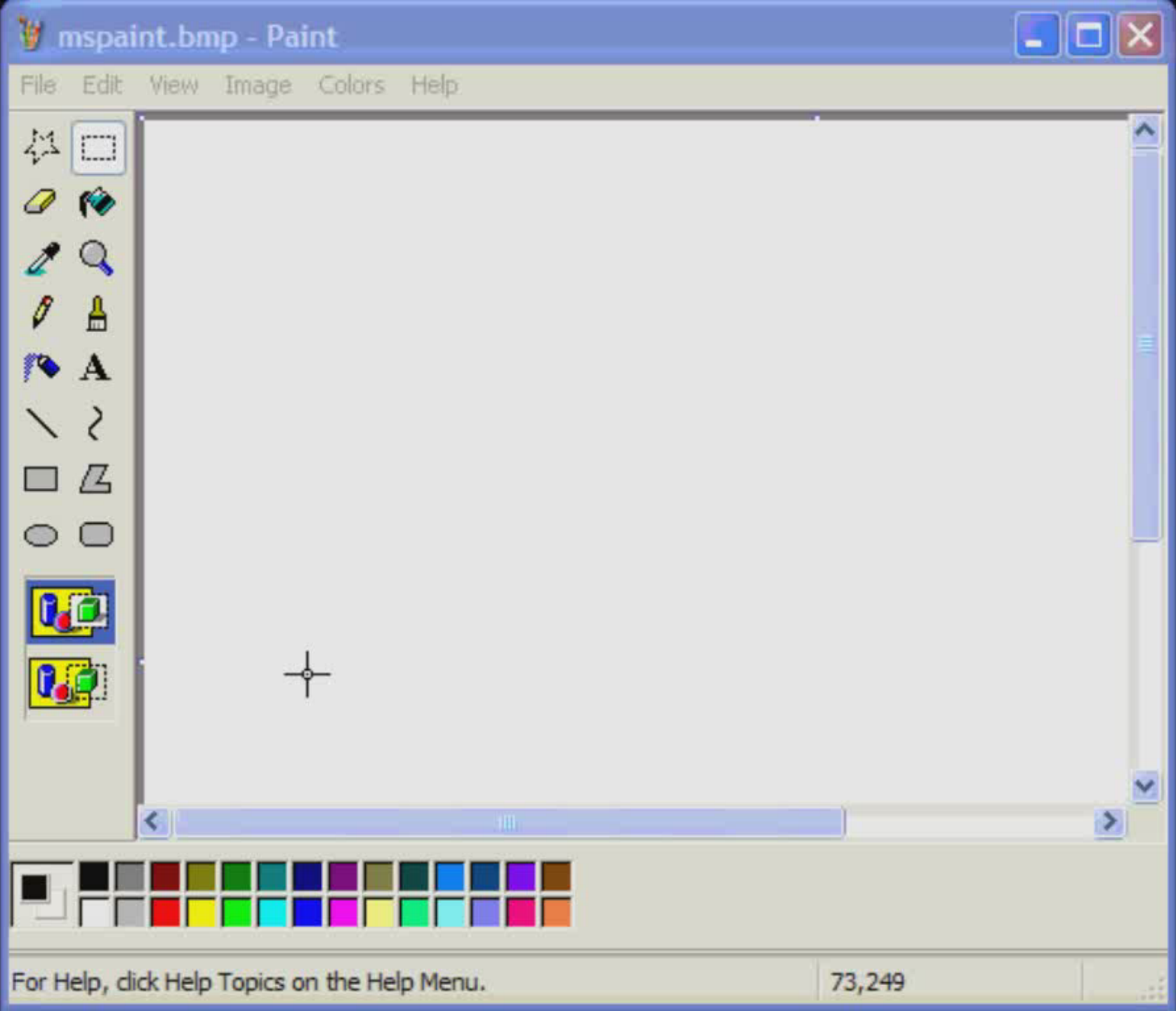


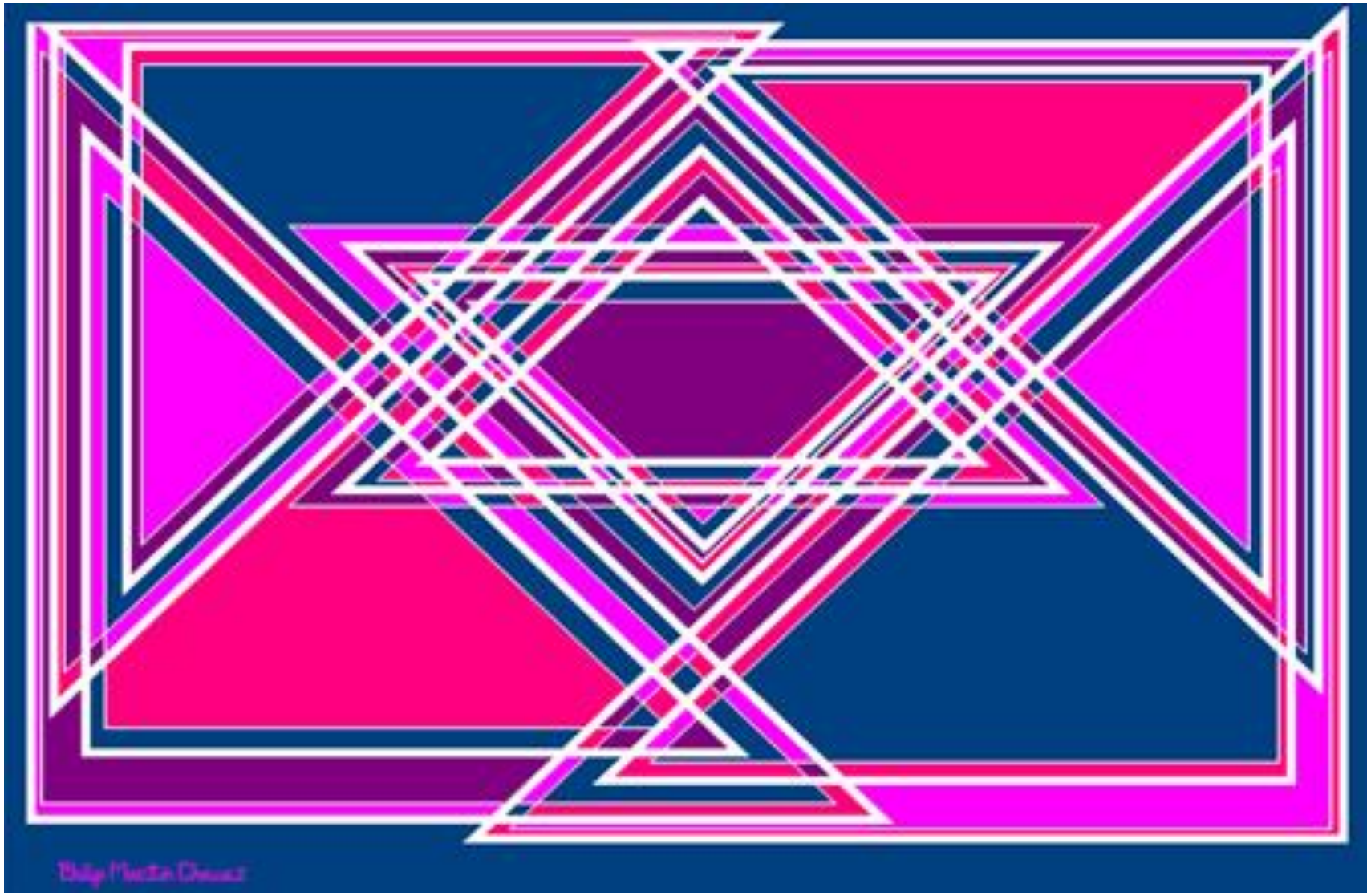
Philip Martin Chavez



Electronic Voice Painter

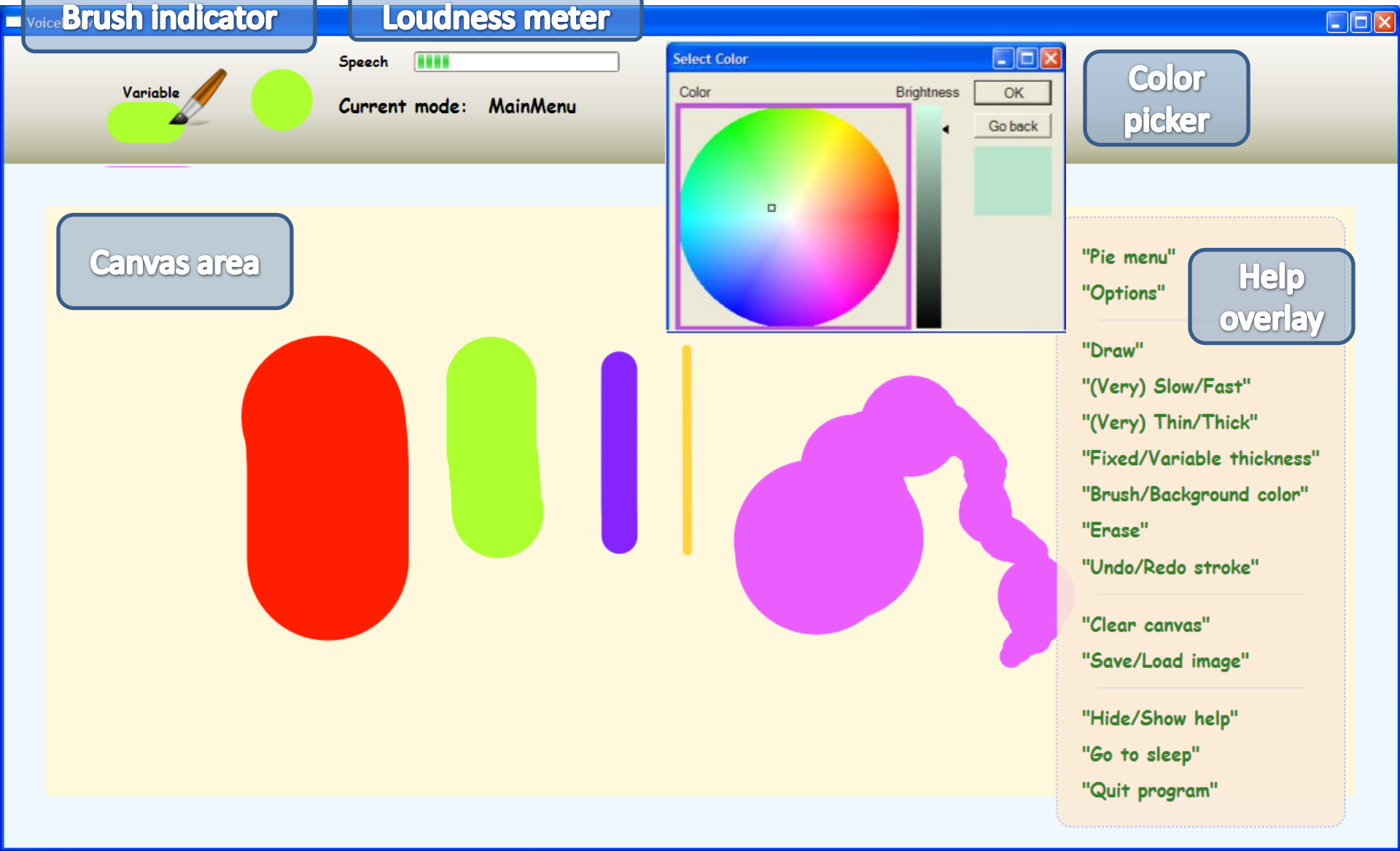

NCA
FUNK


mspaint.bmp - Paint
File Edit View Image Colors Help
For Help, click Help Topics on the Help Menu.
73,249



http://www.voicedraw.com

VoiceDraw



iceDraw

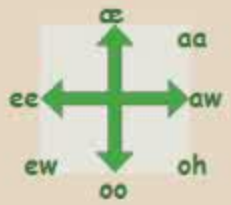
Variable

Speech

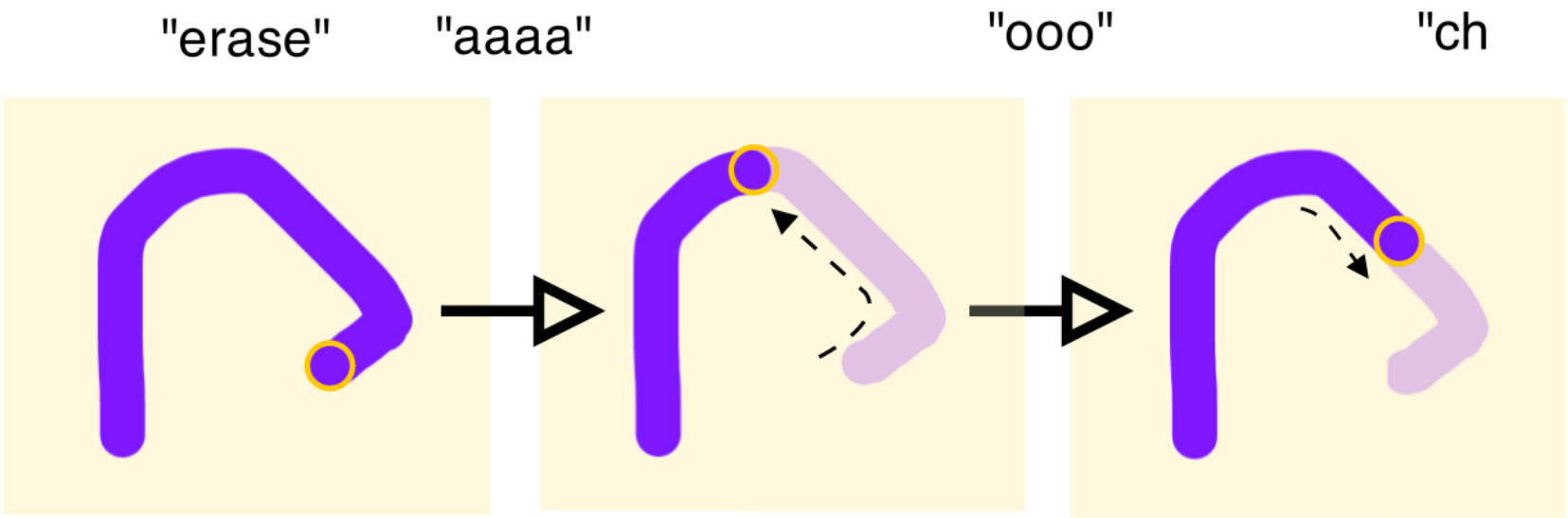
Current mode: BrushUp

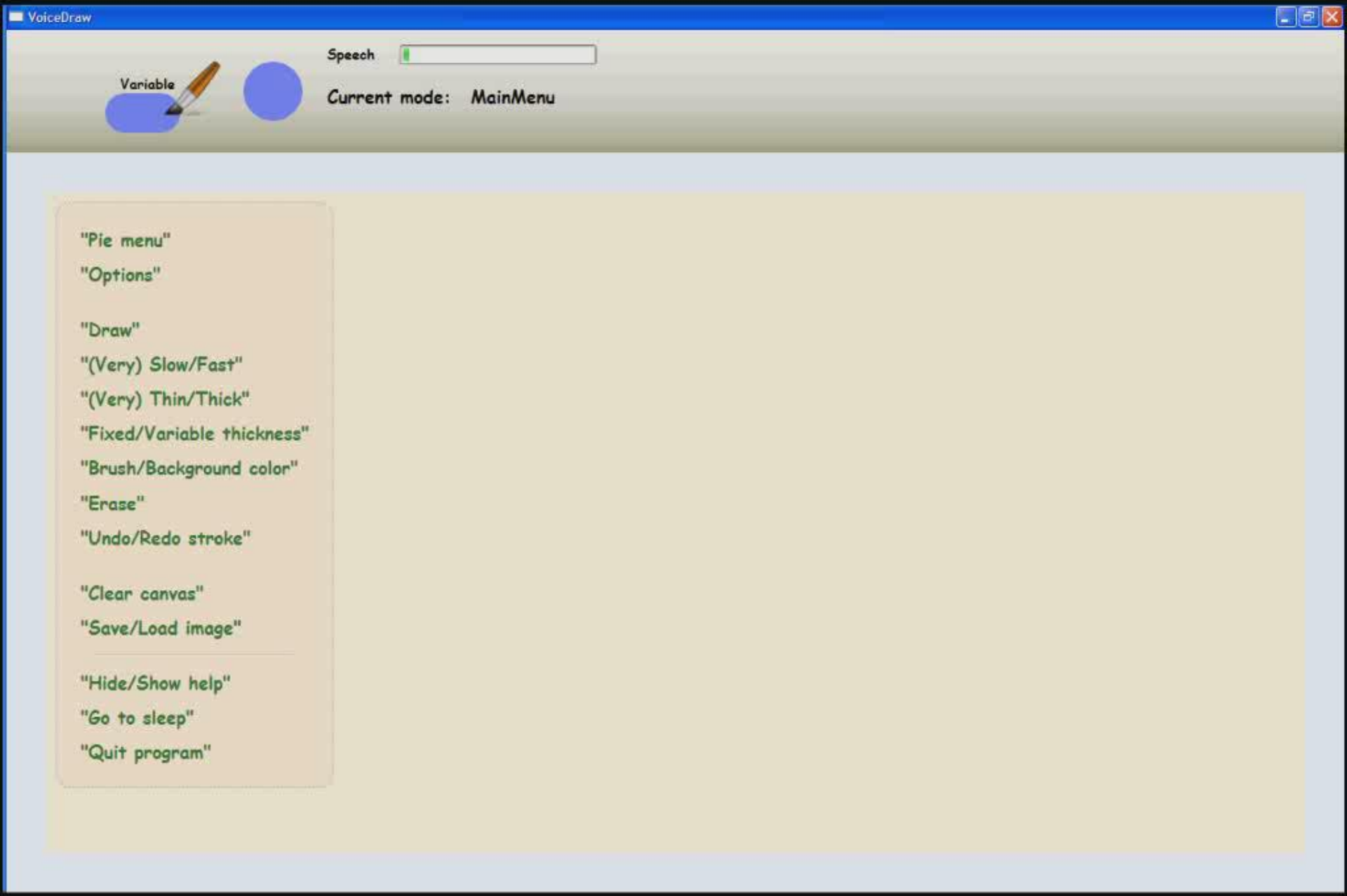
"ck" (Main menu)

"ch" (Brush down)



Continuous undo




VoiceDraw
Variable
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"

VoiceDraw

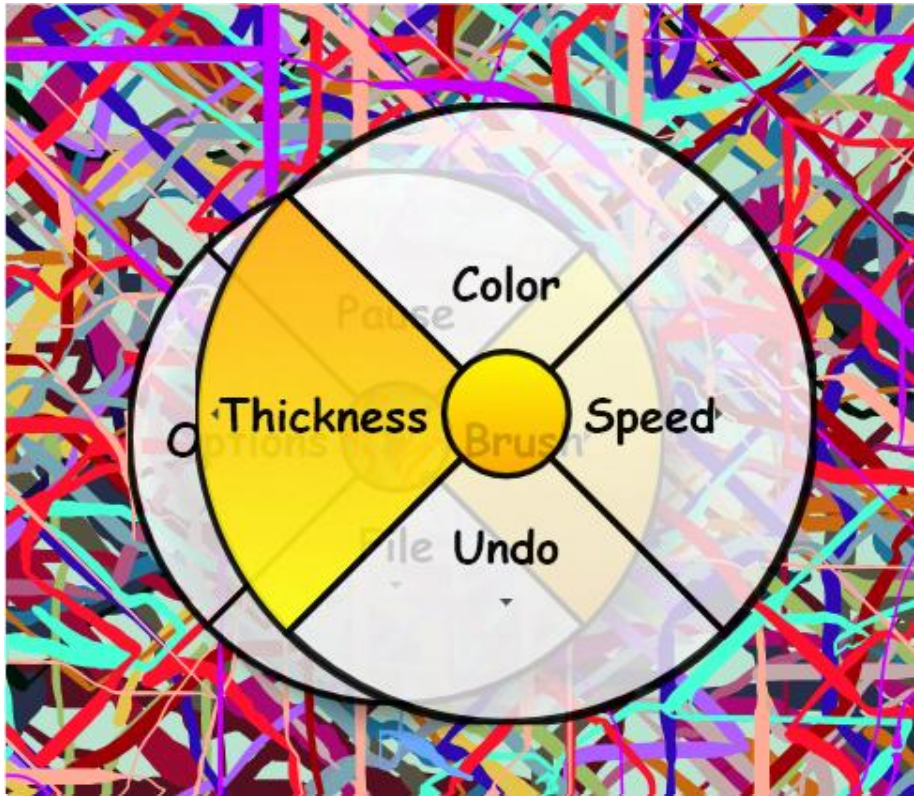
Speech

Current mode: Start

"Wake up"

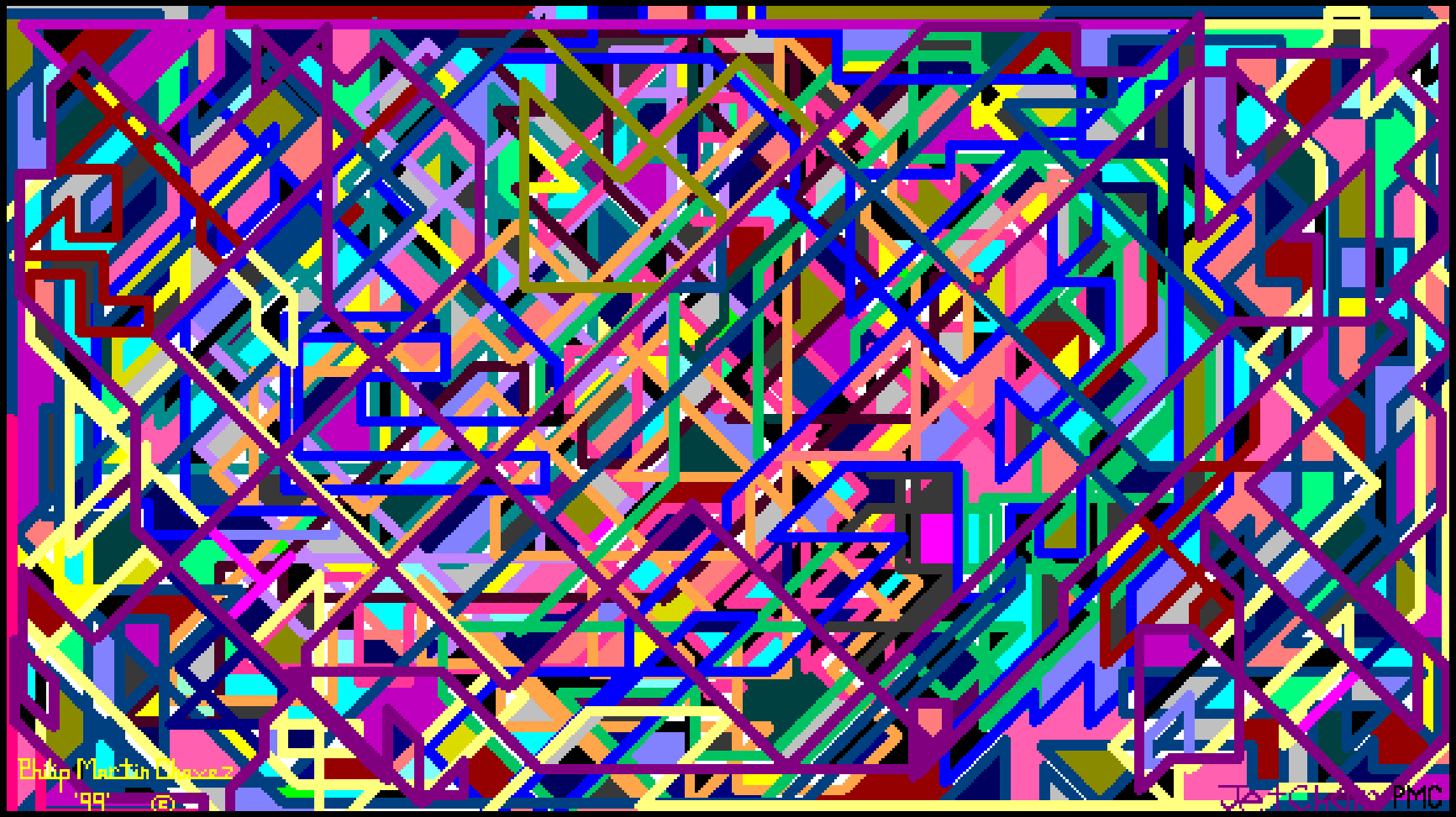
Voice marking menu

“ch” “d”

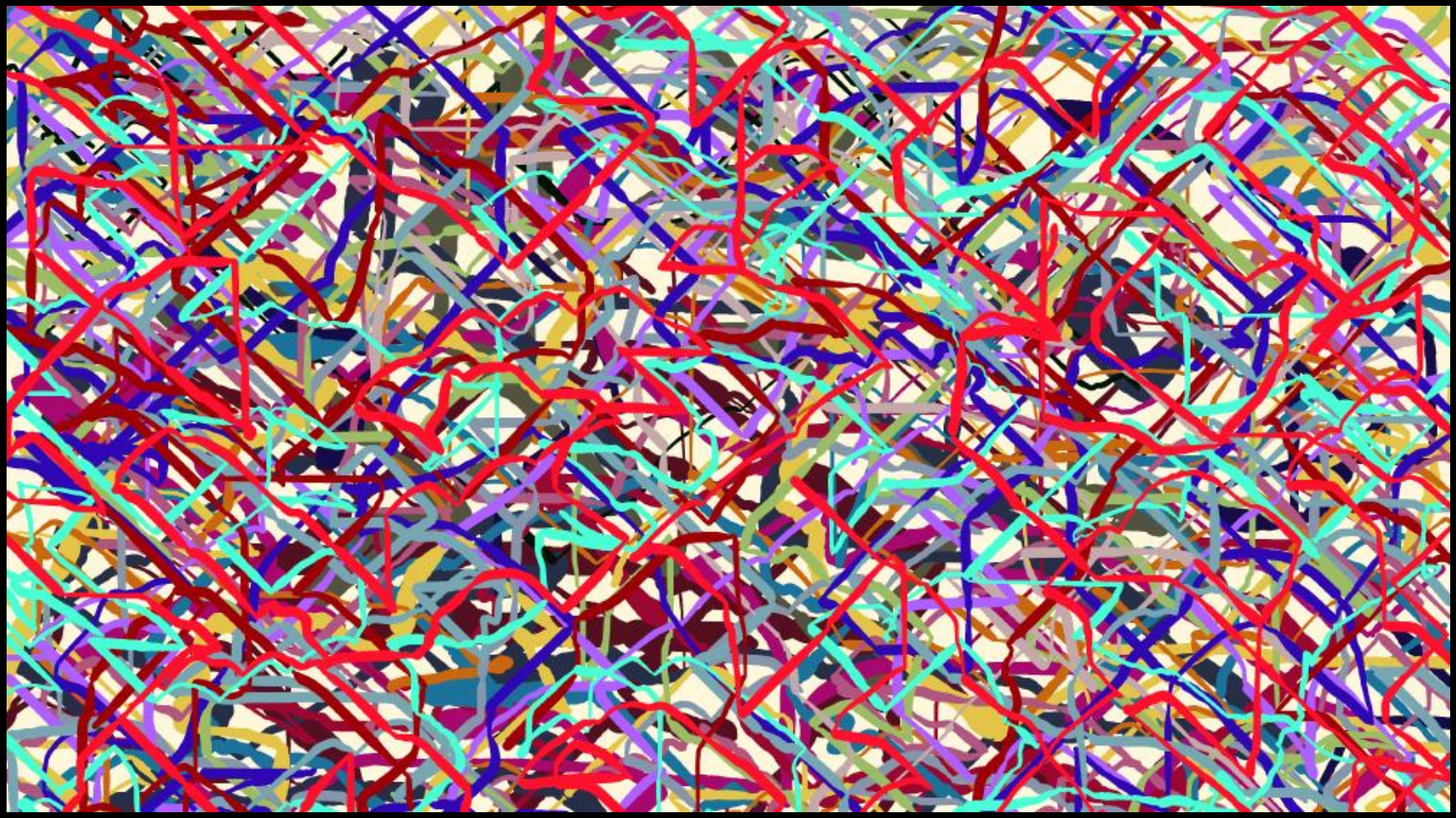


Created using VoiceDraw (2.5 hours)

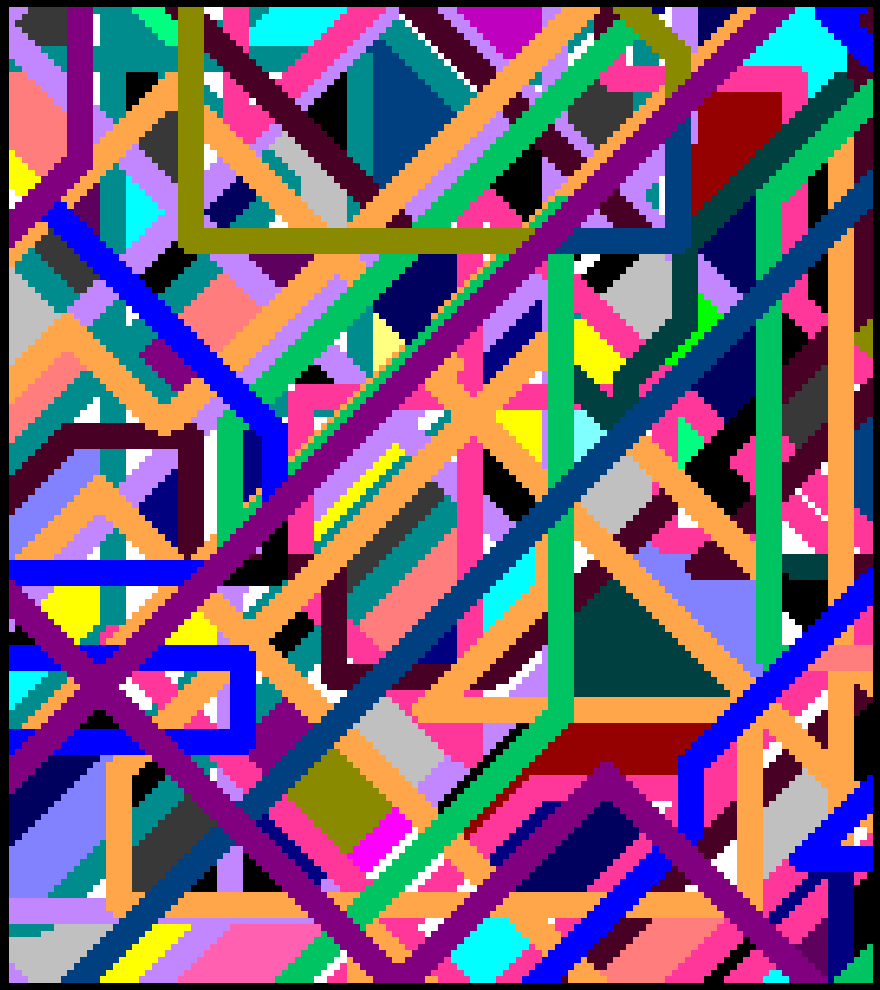
Created using MS Paint & Dragon Dictate (9 hours)



Created using VoiceDraw (3 hours)

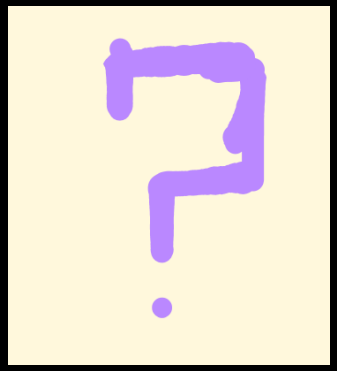


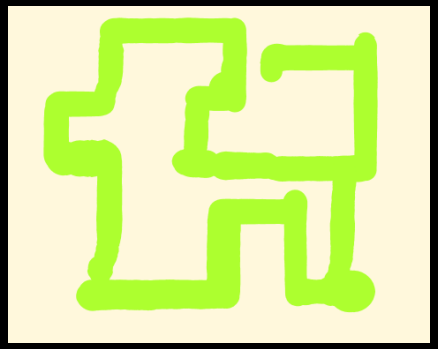
Dragon Dictate & MS Paint

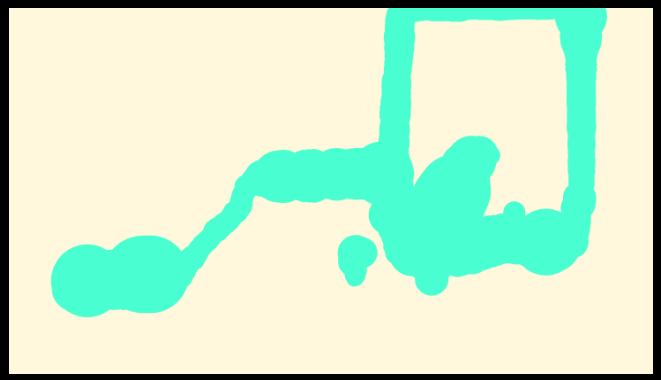


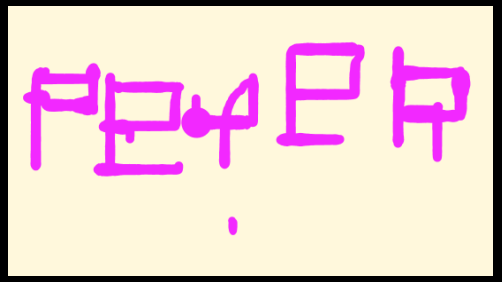
VoiceDraw

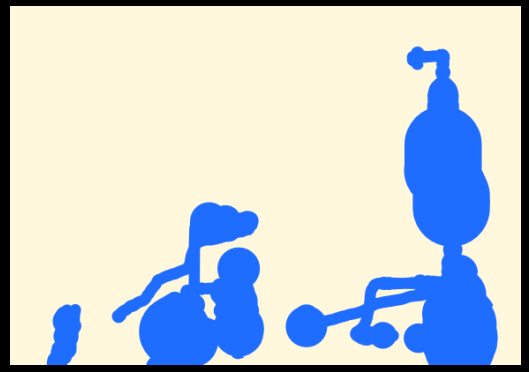

The Vocal Joystick: VoiceBot

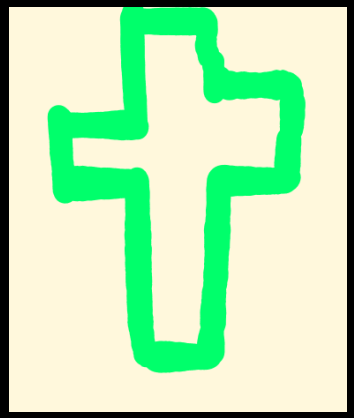


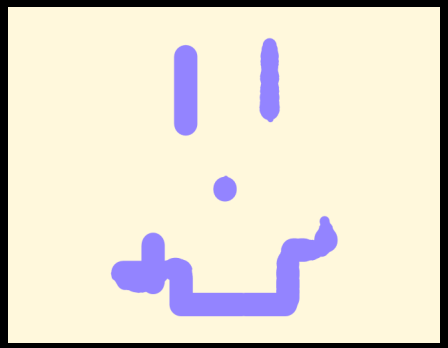




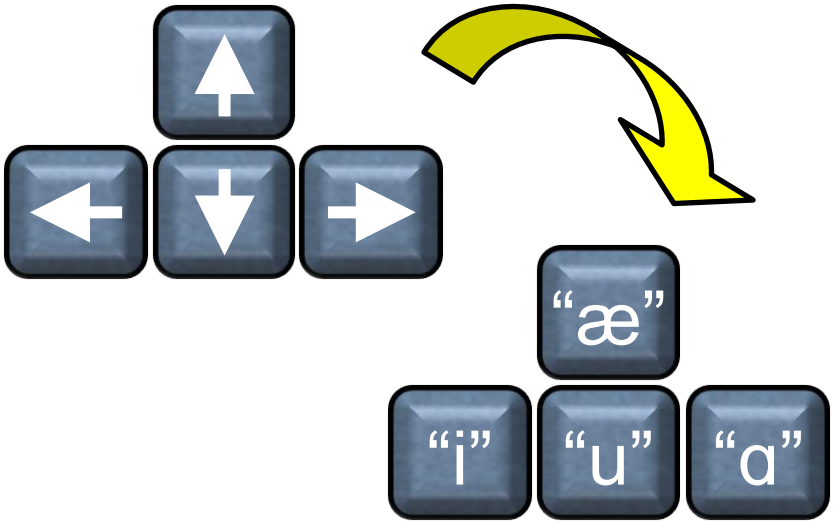


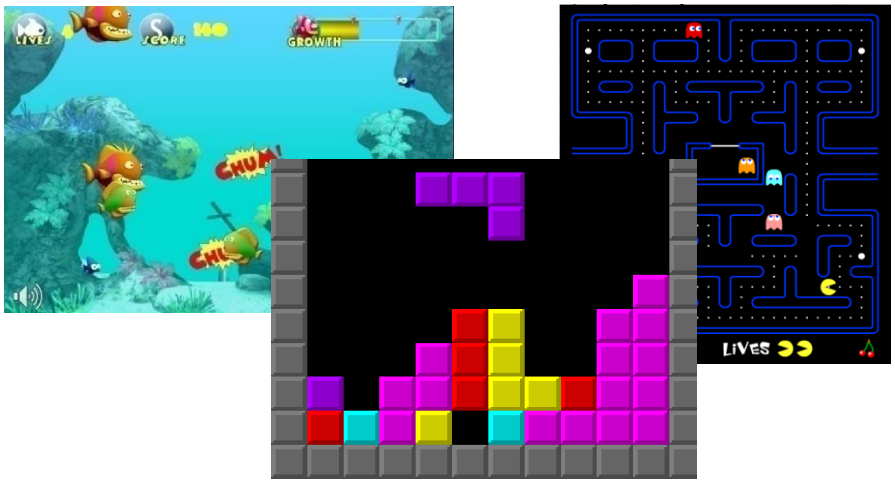


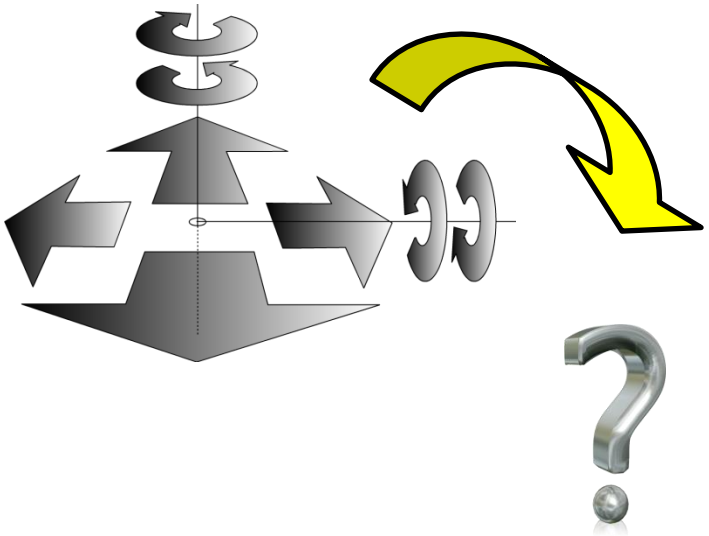




VoiceGames

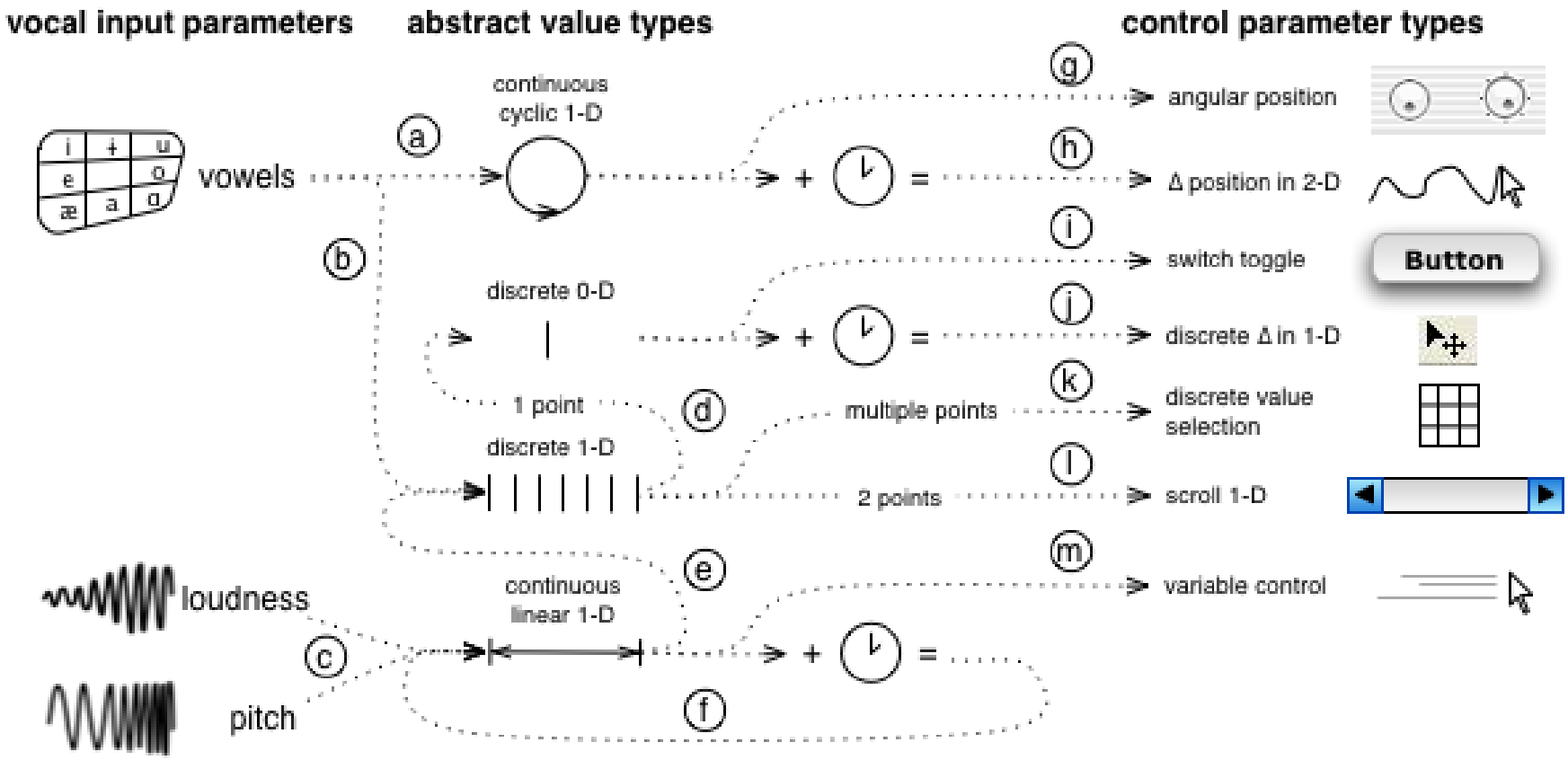

“æ”
“i”
“u”
“a”


LIVES
SCORE
140
GROWTH
CHUA!
CHU!
LIVES




ARMOR: 0
HEALTH: 100
WEAPON: ROCKET LAUNCHER
AMMO: 13/ 60

Vocal parameters



Thank you!
Questions!
Discussions!

Susumu Harada
harada@cs.washington.edu
http://www.vocaljoystick.org