Auditory Display and Interaction

a cognitive engineering perspective

DSP + Computer Music, Michaelmas 2021
Alan Blackwell

Some auditory channel properties

- The cochlea is an FFT machine (so 2D, not 1D)
  - Not just time-domain amplitude samples
  - Instantaneous & continuous frequency band energies
- Auditory mode is closely coupled to language centres
  - Lexical associations
  - (Musical) structure: syntax and rhythm
- Binaural perception provides orientation relative to head

Display modes as information channels

Visual display	Auditory display	Haptic display
World coordinates within restricted field	Approximate head coordinates	Precise contact coordinates
Encoded as image in visuo-spatial sketchpad	Encoded as chunks in phonological loop	Persistent nerve stimulus
Allows direct relational encodings	Allows direct linguistic encodings	Few temporal or semantic conventions
Suffers from occlusion, but not signal masking	Does not require ambient light	Requires physical contact for channel transmission

Auditory display as semiotic system

- What are the perceptual components?
  - Perceived primitives
  - Perceived relations
- What are the application referents?
  - Information entities
  - Information relations
- System design employs modes of correspondence

Does auditory display belong in a computer music course?

- Music is “structured sound”, so yes, because we are defining and exploiting structure through design of correspondences
- Musical analysis offers a useful feature space for synthesis and interpretation, better than trial-and-error, intuition, or first-principles neuroscience:
  - Expressive dynamics
  - Pitch and chroma
  - Harmonic and tonal conventions
  - Timbre (everything else, including modulation envelope, filter design, instrumental technique, resonance and damping …)

Technical varieties of “non-speech audio”

- Alarms and notifications
- Sonification / audialisation
- Spatial navigation
- Audio games / game audio and sound toys
- And some specialised speech variants

Technical varieties of “non-speech audio”

- **Alarms** and notifications
- Sonification / audialisation
- Spatial navigation
- Audio games / game audio and sound toys
- And some specialised speech variants

Alarms

- A classic area of interest in human factors
  - Characteristic of “first-wave HCI” – engineering specification of the human user as a system component
  - Relevant to aircraft cockpits, road safety, control rooms etc.
- Research questions and technical guidance include:
  - What sound properties are distinguishable?
  - What properties gain the user’s attention?
  - What trade-offs exist between urgency and situation assessment?

Beyond the ambulance siren

- Brute force (‘better safe than sorry’) approach, where sounds are used to signal danger or potential danger
- Flooding the environment with sound:
  - (+) Is certain to attract attention
  - (–) Causes startled reactions
  - (–) Prevents communications at a crucial point in time

Patterson's design guidance:

- The number of immediate-action warning sounds should not exceed about six
- Each sound should have a distinct melody and temporal pattern
- It is possible to predict the optimum sound level for a warning sound in most noise environments
- R. D. Patterson (1990). Auditory Warning Sounds in the Work Environment. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 327, No. 1241, Human Factors in Hazardous Situations (Apr. 12, 1990), pp. 485-492

Technical varieties of “non-speech audio”

- Alarms and **notifications**
- Sonification / audialisation
- Spatial navigation
- Audio games / game audio and sound toys
- And some specialised speech variants

Bill Gaver's *Sonic Finder* (1987)

- https://vimeo.com/158610127
- Auditory icons or "earcons"
  - Now a familiar system feature
  - Either unsolicited event notification ("alert")
  - Or status change confirmation (tap, click, etc)
  - Secondary multimodal information channel
- Classic opportunity for semiotic design – what is the mode of correspondence? (Csapó & Wersényi attempt some formal definitions)

  - Csapó, Ádám, and György Wersényi. "Overview of auditory representations in human-machine interfaces." ACM Computing Surveys (CSUR) 46, no. 2 (2013): 19.

Technical varieties of “non-speech audio”

- Alarms and notifications
- **Sonification / audialisation**
- Spatial navigation
- Audio games / game audio and sound toys
- And some specialised speech variants

The design logic of visualisation/sonification

- We have a set of data, and believe there are patterns within that data that we (or the users) are not aware of
- The data is used to generate sensory stimuli, in the expectation that patterns in the data will result in perceptual regularities.
  - In visualisation, these are visual ‘patterns’ (e.g. time series)
  - In sonification, these have to be auditory patterns

Effectiveness of sonification

- The naïve view: audio will be intuitive, e.g. Cowden project
- The design view: the expected relations must be designed
  - Program code – despite various attempts, this hasn't worked yet
  - Scientific/financial/media data – might at least entertain:
    - e.g. https://www.youtube.com/watch?v=0lKSFlB_-Q0
  - Status and alerts (surgical 'bing', EDSAC) – see Patterson
  - Structural coupling – e.g. filling a wine glass?
- As with visualisation, correspondences must be designed

Technical varieties of “non-speech audio”

- Alarms and notifications
- Sonification / audialisation
- **Spatial navigation**
- Audio games / game audio and sound toys
- And some specialised speech variants

Real-world navigation

- Mcgookin, Brewster & Priego (2009). Audio bubbles: Employing non-speech audio to support tourist wayfinding. Int. Conf. Haptic and Audio Interaction Design, pp. 41-50
- Wilson, Walker, Lindsay, Cambias & Dellaert (2007). Swan: System for wearable audio navigation. IEEE Int. Symp. Wearable Computers, pp. 91-98

Virtual-world navigation

- In Menus: Brewster (1998). Using nonspeech sounds to provide navigation cues. ACM Trans. on CHI, 5(3), 224-259.
  - Earcons indicate position within hierarchy
- In Games: Papa Sangre II
  - Follow direction cues and sound effects without visuals

Technical varieties of “non-speech audio”

- Alarms and notifications
- Sonification / audialisation
- Spatial navigation
- **Audio games / game audio and sound toys**
- And some specialised speech variants

‘Sound toys’ as interactive compositions

- Dolphin, A. (2014). Defining sound toys: play as composition. In *The Oxford Handbook of Interactive Audio*.
  - Interactive, sonic-centric systems in which the user may trigger, generate, modify, or transform sound.
  - Playful approaches to novice composition through symbolic representation of complex underlying systems.
- e.g. Eno & Chilvers *Bloom*, Bjork *Biophilia*, Coldcut *Ninja Jamm*

Technical varieties of “non-speech audio”

- Alarms and notifications
- Sonification / audialisation
- Spatial navigation
- Audio games / game audio and sound toys
- **And some specialised speech variants**

Using recognisable sounds

- Spearcons (earcons with speech): compressed speech extracts with pitch preserved
- Spindex (speech index): audio with letter sounds or words when scrolling through an index
- Musicons: Recognisable short extracts of music
  - leitmotifs in Wagner (e.g. Nothung) https://pjb.com.au/mus/wagner/index.html
  - samples as signifiers (e.g. James Brown)

Some researchers to follow

- Bruce Walker at Georgia Tech
  - Sonification and Spearcons
- Tony Stockman at Queen Mary
  - Especially for assistive technology
- Steve Brewster in Glasgow
  - Especially for haptic and multimodal mobile