Multi-Sensory Systems

Overview

Multi-sensory systems use more than one sensory channel in interaction

E.g. sounds, text, hypertext, animation, video, gestures, vision etc.

Used in a range of applications: particularly good for users with special needs, and virtual reality

Topic Summary

We will cover

- general terminology
- speech
- non-speech sounds
- handwriting
- text and hypertext
- animation and video

considering applications as well as principles

Usable Senses

The 5 senses (sight, sound, touch, taste and smell) are used by us every day

- each is important on its own
- together, they provide a fuller interaction with the natural world

Computers rarely offer such a rich interaction

Can we use all the available senses?

ideally, yes

practically - no

We can use

- sight
- sound
- touch (sometimes)

We cannot (yet) use

- taste
- smell

Multi-modal versus Multi-media

Multi-modal systems

- use more than one sense (or *mode*) of interaction

  - e.g. visual and aural senses: a text processor may speak the words as well as echoing them to the screen

Multi-media systems

- use a number of different media to communicate information

- e.g. a computer-based teaching system may use video, animation, text and still images: different media all using the visual mode of interaction. It may also use sounds, both speech and non-speech: two more media, now using a different mode.

Speech

Human beings have a great and natural mastery of speech

- makes it difficult to appreciate the complexities, but
- it’s an easy medium for communication

Structure of Speech

- *phonemes* - 40 of them: basic atomic units, which sound slightly different depending on the context they are in; this larger set of sounds are

- *allophones* - all the sounds in the language: between 120 and 130 of them. These are formed into

- *morphemes* - smallest unit of language that has meaning.

Other terminology:

- *prosody* - alteration in tone and quality: allows variations in emphasis, stress, pauses and pitch to impart more meaning to sentences.
- *co-articulation* - the effect of context on the sound; co-articulation transforms the set of phonemes into the set of allophones.
- *syntax* - structure of sentences
- *semantics* - meaning of sentences

Speech Recognition Problems

Different people speak differently: accent, intonation, stress, idiom, volume and so on can all vary.

The syntax of semantically similar sentences may vary.

Background noises can interfere.

People often “ummm.....” and “errr.....”

Recognising words is not the ultimate goal of a speech recognition system: the semantics have to be extracted as well. It often requires intelligence to understand a sentence: the context of the utterance often has to be known, as does information about the subject and sometimes the speaker.

Example:

Even if

“Errr.... I, um, don’t like this”

is recognised, it is a fairly useless piece of information on it’s own

The Phonetic Typewriter

Developed for Finnish (a phonetic language, written as it is said).

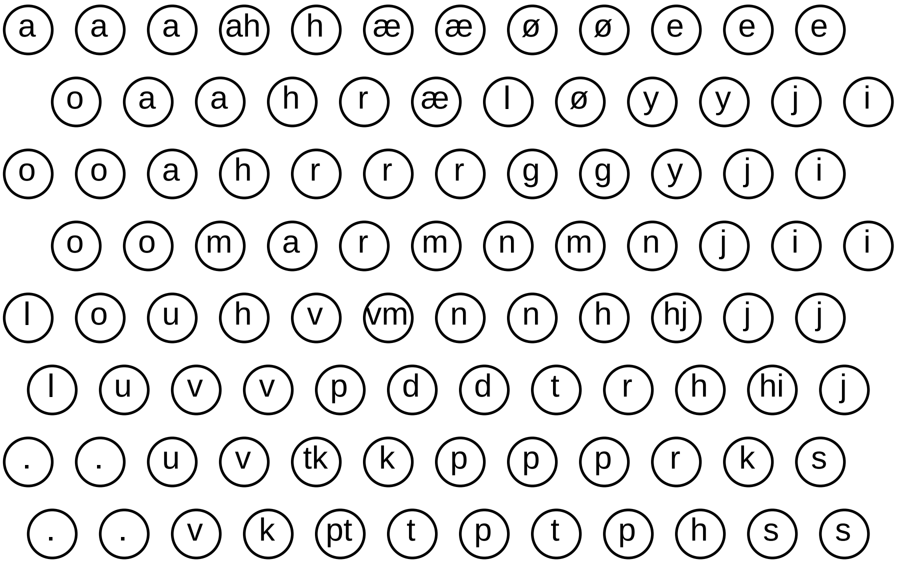
Trained on one speaker, will generalise to others.

A neural network is trained to cluster together similar sounds, which are then labelled with the corresponding character.

When recognising speech, the sounds uttered are allocated to the closest corresponding output, and the character for that output is printed.

• requires large dictionary of minor variations to correct general mechanism

• noticeably poorer performance on speakers it has not been trained on



a a a ah h æ æ ø ø e e e
o a a h r æ l ø y y j i
o o a h r r r g g y j i
o o m a r m n m n j i i
l o u h v vm n n h hj j j
l u v v p d d t r h hi j
. . u v tk k p p p r k s
. . v k pt t p t p h s s

Speech Recognition: currently useful?

Single user, limited vocabulary systems can work satisfactorily

No general user, general vocabulary systems are commercially successful, yet

Large potential, however

- when users hands are already occupied - manufacturing, for example
- for users with physical disabilities
- lightweight, mobile devices

Speech Synthesis

Speech synthesis: the generation of speech

Useful - natural and familiar way of receiving information

Problems - similar to recognition: prosody particularly

Additional problems

• intrusive - either requires headphones, or creates noise in the workplace

• transient - harder to review and browse

Successful in certain constrained applications, usually when the user is particularly motivated to overcome the problems and has few alternatives

• screen readers - read the textual display to the user: utilised by visually impaired people

• warning signals - spoken information is sometimes presented to pilots whose visual and haptic skills are already fully occupied

Non-Speech Sounds

Boings, bangs, squeaks, clicks etc.

- commonly used in interfaces to provide warnings and alarms

Evidence to show they are useful

- fewer typing mistakes with key clicks

- video games harder without sound

Dual mode displays: information presented along two different sensory channels

Allows for redundant presentation of information - the user can utilise whichever they find easiest

Allows resolution of ambiguity in one mode through information contained in the other

Sound especially good for transient information, and background status information

It is also language/culture independent, unlike speech

Example: Sound can be used as a redundant mode in the Apple Macintosh; almost any user action (file selection, window active, disk insert, search error, copy complete, etc.) can have a different sound associated with it.

Auditory Icons

Use natural sounds to represent different types of object or action

Natural sounds have associated semantics which can be mapped onto similar meanings in the interaction

- e.g. throwing something away can be represented by the sound of something smashing

Problem: not all things have associated meanings: e.g. copying

Application: SonicFinder for the Macintosh

Items and actions on the desktop have associated sounds

- folders have a papery noise
- moving files is accompanied by a dragging sound
- copying (a problem one) has the sound of a liquid being poured into a receptacle; the rising pitch indicates the progress of the copy
- big files have a louder sound than smaller ones

Additional information can also be presented:

- muffled sounds indicate the object is obscured or an action is in the background
- use of stereo allows positional information to be added

Earcons

Synthetic sounds used to convey information

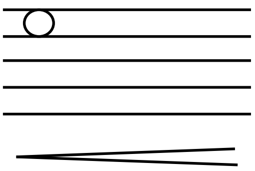
Structured combinations of notes, called *motives* , used to represent actions and objects

Motives combined to provide rich information

• compound earcons
multiple motives combined to make one more complicated earcon: for example

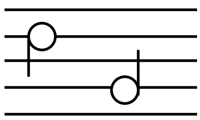
Create

note, getting louder



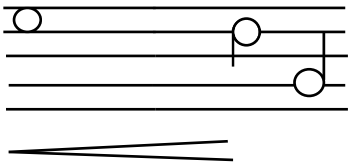
File

high-low note



Create file

create icon followed by file icon



• family earcons
similar types of earcons represent similar classes of action or similar objects: the family of “errors” would contain syntax and operating system errors

Earcons easily grouped and refined due to compositional and hierarchical nature

Harder to associate with the interface task since there is no natural mapping

Handwriting recognition

Handwriting is another communication mechanism which we are used to

Technology

Handwriting consists of complex strokes and spaces

Captured by digitising tablet - strokes transformed to sequence of dots

- large-scale tablets available, more suitable for digitising maps and technical drawings
- smaller devices, some incorporating thin screens to display the information, becoming available e.g. those produced by Apple as personal organisers

Recognition

Problems

- personal differences in letter formation
- co-articulation effects

Of limited success are systems that are trained on a few users, with separated letters

Generic multi-user naturally-written text recognition systems are not currently of significant accuracy to be commercially successful

Text and Hypertext

Text is a common form of output, and very useful in many situations

• imposes a strict linear progression on the reader, according to the author's ideas of what is best - this may not be ideal

Hypertext structures blocks of text into a mesh or network that can be traversed in many different ways

• allows a user to follow their own ideas and concepts through information

• hypertext systems comprise:

  - a number of pages, and
  - links, that allow one page to be accessed from another

Example

A technical manual for a photocopier may have all the technical words linked to their definition in a glossary. It may be possible to follows links so that one reads all the information on a particular aspect of the system, such as all the electronics, or to follow a different route through the data to solve a problem with, say, the copying to double-sided documents. Many of the pages visited will be identical in both cases, but will be encountered in a different order

Hypermedia

Hypermedia systems are hypertext systems that incorporate additional media, such as illustrations, photographs, video and sound

Particularly useful for educational purposes

• animation and graphics can allow user to see things happen as well as read

• hypertextual structure allows users to explore at their own pace following threads that interest them

Problems

• “lost in hyperspace” - users can be unsure as to where in the hypertext web they are
Maps of the hypertext are a partial solution, but since hypertexts can be large these can be daunting too

• incomplete coverage of information
As there are so many different routes through the hypertext, it is possible to miss out chunks, by taking routes that avoid these areas

• difficult to print out and take away
Printed documents require a linear structure; it can be difficult to get the relevant information printed out in a neat manner

Animation

Animation refers to the addition of motion to images; they change and move in time

Simple examples:

- clocks
  - Digital faces - seconds flick past
  - Analogue face - second hand sweeps round constantly
  - Salvador Dali clock - digital numbers warp and melt, one digit into the next
- cursor
  - hourglass/watch/spinning disc indicates the system is busy
  - flashing cursor indicates typing position clearly
  - different types of cursor pointer indicate different functionality available, or different mode

Animation used to great effect to indicate temporally-varying information.

Useful in education and training: allow users to see things happening, as well as being interesting and entertaining images in their own right

Example: data visualisation

Abrupt and smooth changes in multi-dimensional data can be visualised using animated, coloured surfaces that ripple and fluctuate.

Complex molecules and their interactions can be more easily understood when they are drawn and moved on the screen, rotated and viewed from arbitrary positions.

Video and Digital Video

Compact disc technology is revolutionizing multimedia systems: large amounts of video, graphics, sound and text can be stored and easily retrieved on a relatively cheap and accessible medium

Different approaches, characterised by different compression techniques that allow more data to be squeezed onto the disc

- CD-I: excellent for full-screen work. Limited video and still image capability; targeted at domestic market

- CD-XA (eXtended Architecture): development of CD-I, better digital audio and still images

- DVI (Digital Video Interactive)/UVC (Universal Video Communications): support full motion video

Example: *Palenque* - a DVI-based system

Multimodal multimedia prototype system, in which users wander around a Mayan site. Uses video, images, text and sounds.

QuickTime from Apple represents a standard for incorporating video into the interface. Compression, storage, format and synchronisation are all defined, allowing many different applications to incorporate video in a consistent manner.

Utilising animation and video

Animation and video are potentially powerful tools

- notice the success of television and arcade games

However, the standard approaches to interface design do not take into account the full possibilities of such media

We will probably only start to reap the full benefit from this technology when we have much more experience.

We also need to learn from the masters of this new art form: interface designers will need to acquire the skills of film makers and cartoonists as well as artists and writers.

Applications

Users with special needs have specialised requirements which are often well-served by multimedia and/or multimodal systems

- visual impairment - screen readers, SonicFinder
- physical disability - speech input, gesture recognition, predictive systems (e.g. Reactive keyboard)
- learning disabilities (e.g. dyslexia) - speech input, output

Virtual Reality

Multimedia multimodal interaction at its most extreme, VR is the computer simulation of a world in which the user is immersed.

- headsets allow user to "see" the virtual world
- gesture recognition achieved with DataGlove (lycra glove with optical sensors that measure hand and finger positions)
- Eyegaze allows users to indicate direction with eyes alone

Examples:

VR in chemistry - users can manipulate molecules in space, turning them and trying to fit different ones together to understand the nature of reactions and bonding
Flight simulators - screens show the "world" outside, whilst cockpit controls are faithfully reproduced inside a hydraulically-animated box