GuineaPig –
A generic subjective test system for multichannel audio

guinea pig n

   1: a small stout–bodied short–eared nearly tailless domesticated rodent *(Cavia cobaya)* often kept as a pet and widely used in biological research **2:** a subject of scientific research, experimentation, or testing

   –Webster

Jussi Hynninen

Laboratory of Acoustics and Audio Signal Processing
Helsinki University of Technology
Espoo, Finland

`hynde@acoustics.hut.fi`
`www.acoustics.hut.fi/~hynde/`

Nick Zacharov

Speech and Audio Systems Laboratory
Nokia Research Center
Tampere, Finland

`nick.zacharov@research.nokia.com`

GuineaPig – Background / Motivation

Subjective testing is an arduous task

- **Inefficient to set–up and perform**

- **Prone to human error**

- **Difficult to automate**

- **Often requires specific hardware**
  (multichannel I/O, specific UI)

GuineaPig – Background / Motivation

Robust testing requires specific tools

- **Generic access to different tests**
  (appropriate psychometric procedures)

- **Flexibility to customise test**
  (scales, questions, UI, etc.)

- **Fast test generation**
  (for commercial application)

- **Access to block designs**

- **Robust data logging & easy data transfer for analysis**

GuineaPig – Background / Motivation

Some other methods/systems

- **Manual** *(time consuming & error prone)*

- **User programmed** *(time consuming)*
  - **non–trivial and time consuming to implement**
  - **often over–simplified**

- **Designed to perform single tests** *(specific)* **A/B/X, triple stimulus hidden reference, etc.**

- **Tied to other hardware** *(expensive)* **workstation slaving AV playback equipment**

GuineaPig – Summary

- **flexible platform for generic audio (visual) subjective testing**

- **completely hard disk / software based**

- **scalable test platform**

    - **digital / analog audio output**
    - **1 – 8 channels** *(more in near future)*
    - **8 – 48kHz sampling rates**
    - **1 – N GUI grading terminals**

- **provides a wide range of test paradigms**

- **can be extensively and easily customised by the experimenter**

- **runs on SGI IRIX system, written in Java & C**

GuineaPig – Tests

- **Several test types possible:**

    - **single stimulus**
    - **A/B**
    - **A/B/X**
    - **A/B/Ref**
    - **A/B scale**
    - **A/B scale, fixed and hidden reference**
    - **TAFC**
    - **Rank–order**

- **In general: N (1–3 usually) samples are compared / graded**

GuineaPig – Tests

- **Test items define which samples are compared in each case. For example, an A/B test:**

an items–file

```
item1.A:        pirr44
item1.B:        pirr32

item2.A:        pirr22
item2.B:        pirr32

item3.A:        pirr8
item3.B:        pirr11
```

IDs of samples to compare

- **Playlists: define order of (a subset of) items that are presented to subject**

a playlist–file

```
item1
item3
item2
```

GuineaPig – Tests

- **Sample playback:**

  - **Fixed playback sequence, configurable with optional pauses between samples**

  - **Free playback, subject plays samples as he/she wants**

  - **Sample switch using a cross–fade** *(free playback only)*

- **Answering time limit** *(optional)*

- **MCLL (most comfortable listening level): fixed or subject selects at test start**
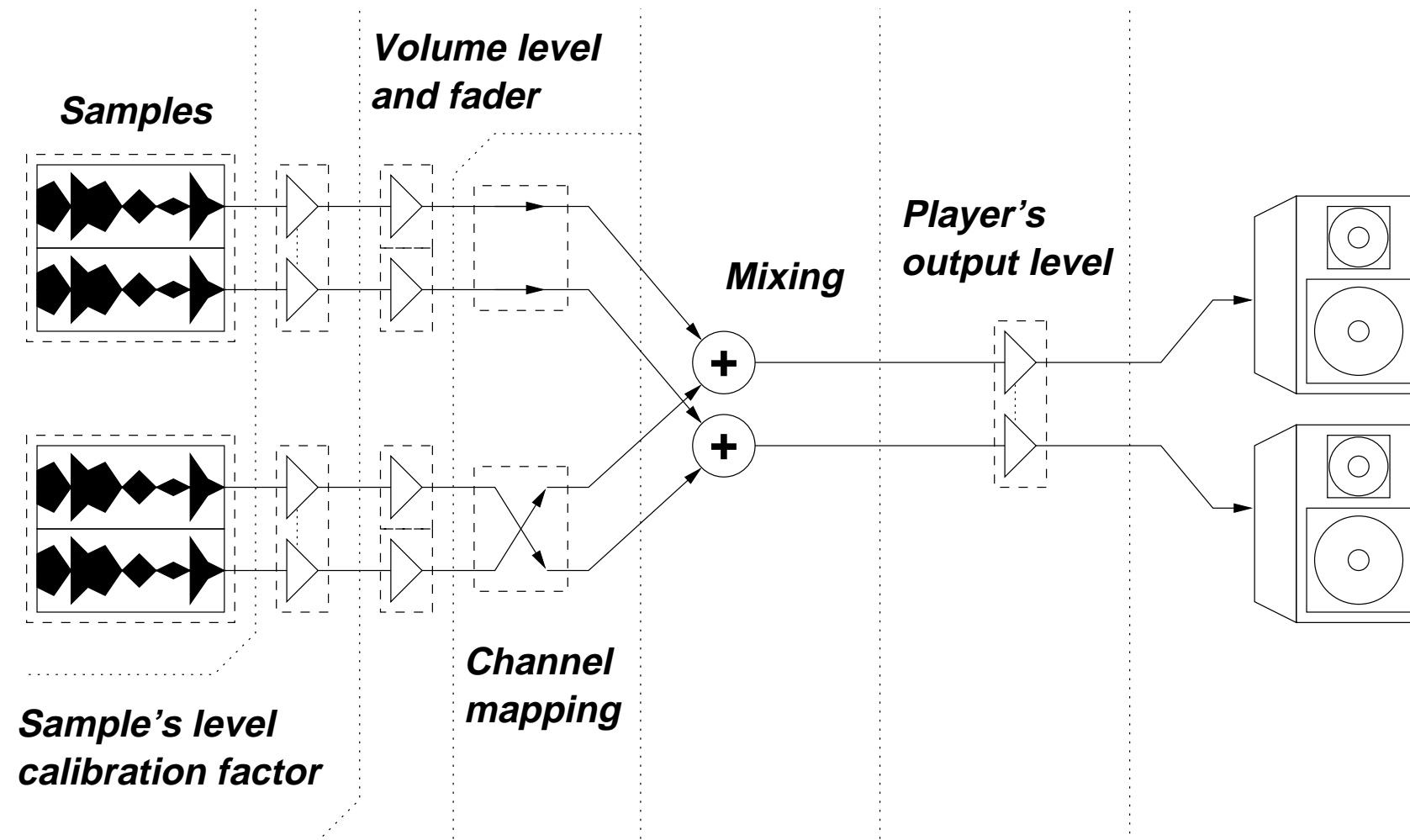
GuineaPig – Audio features

- **Uses SGI's *Audio*, *Audio File*, and *Digital Media* libraries**

- **Sample rates: 8kHz – 48kHz**

- **Outputs options: analog stereo, digital outputs** *(ADAT optical, SP/DIF, AES3)*

- **24–bit output, sound processed in floating point**

- **1–8 chan. output (8 with ADAT), in future more with multiple ADAT interfaces**

- **Audio file formats: AIFF/AIFF–C, WAV, MPEG1, etc.**
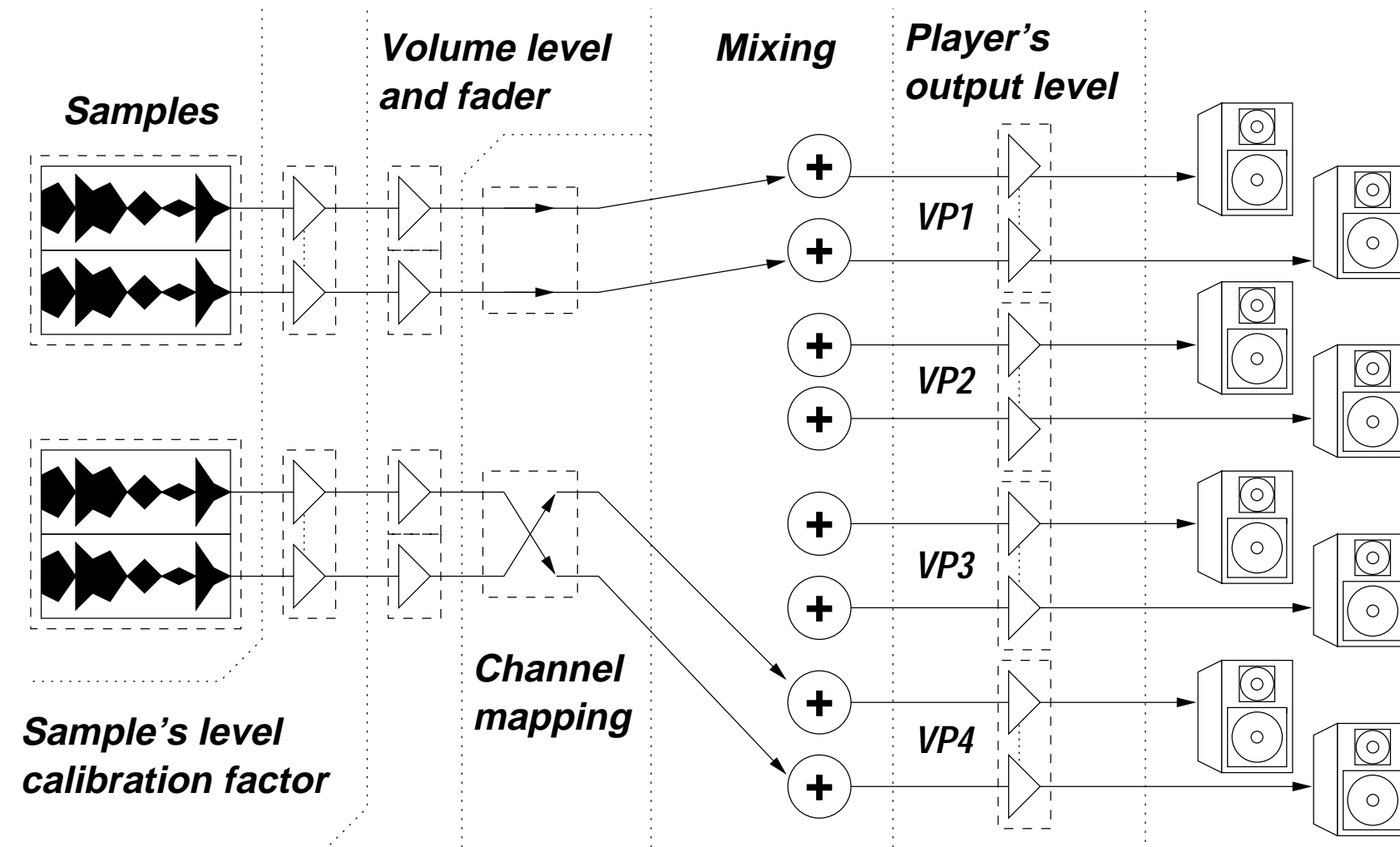
GuineaPig – Audio features

Sound player:

- **Plays multiple samples from hard disk, mixing them together**

- **Volume controls: output, sample volume (with faders), sample level "calibration factor", levels can be given in *linear*, *decibel*, *%–scales***

- **Cross–fades**

- **Delay / latency can be measured and adjusted**

- **Detects audio drop–outs**

- **Synchronyous operations, sample start, fades start at the same time for multiple samples**

GuineaPig – Audio features

Samples

Volume level and fader

Player's output level

Mixing

Channel mapping

Sample's level calibration factor
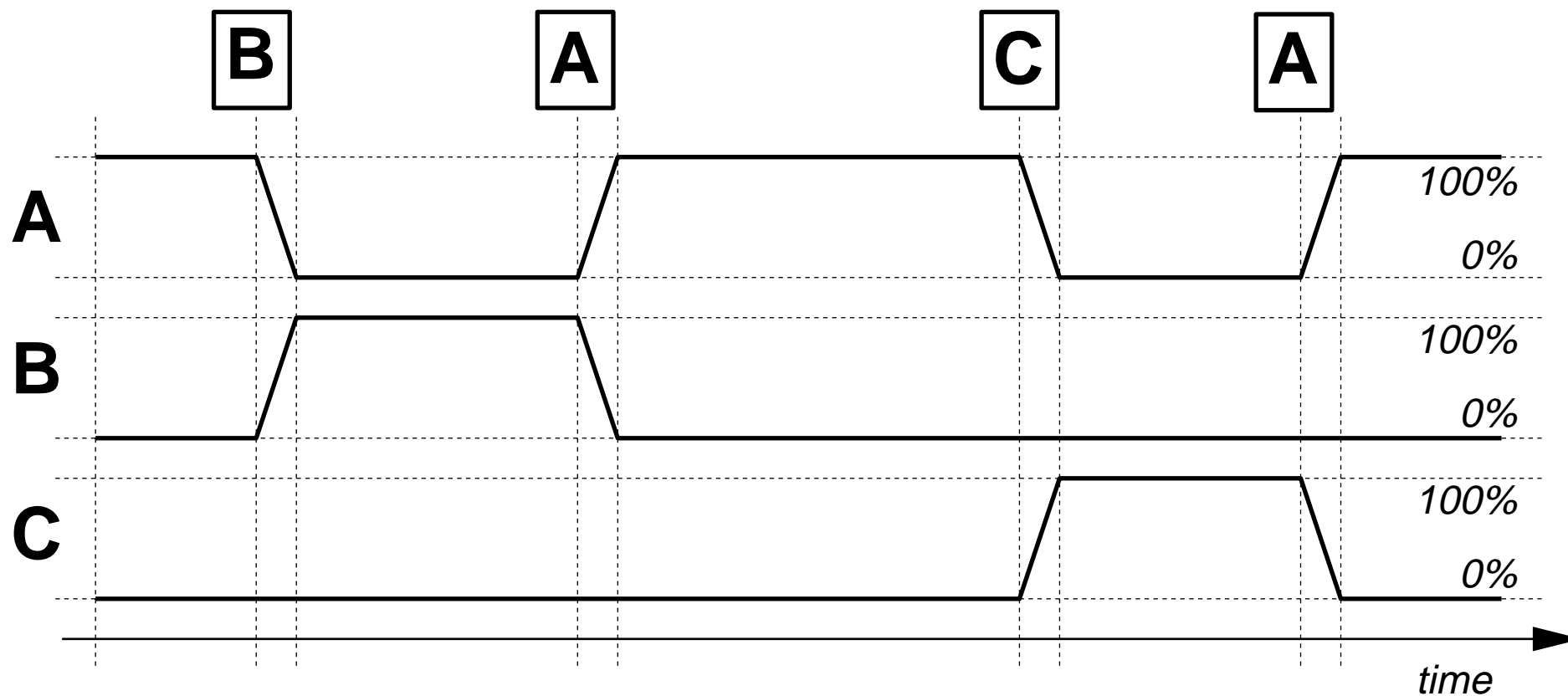
GuineaPig – Audio features

- **Virtual players*: divide output into smaller parts, ex.: 8–ch. output into four stereo–outputs**

* Not fully integrated to testing yet
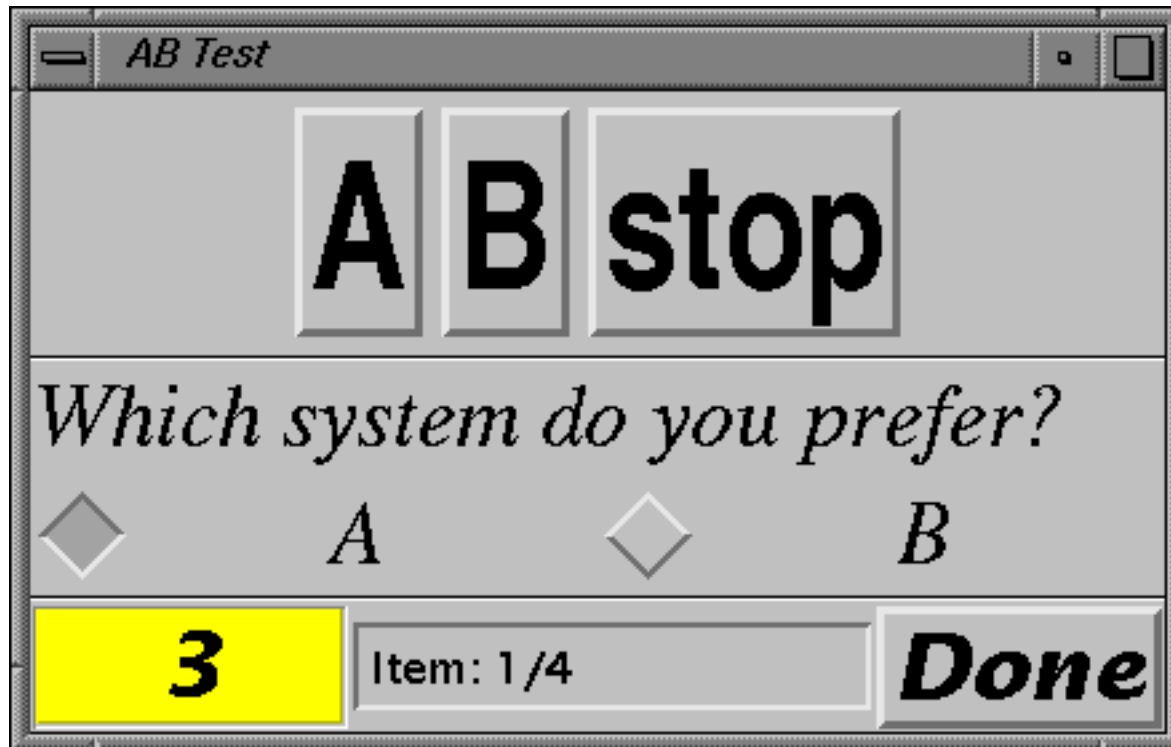
GuineaPig – Audio features

Cross–fades

- **Length of fade configurable**
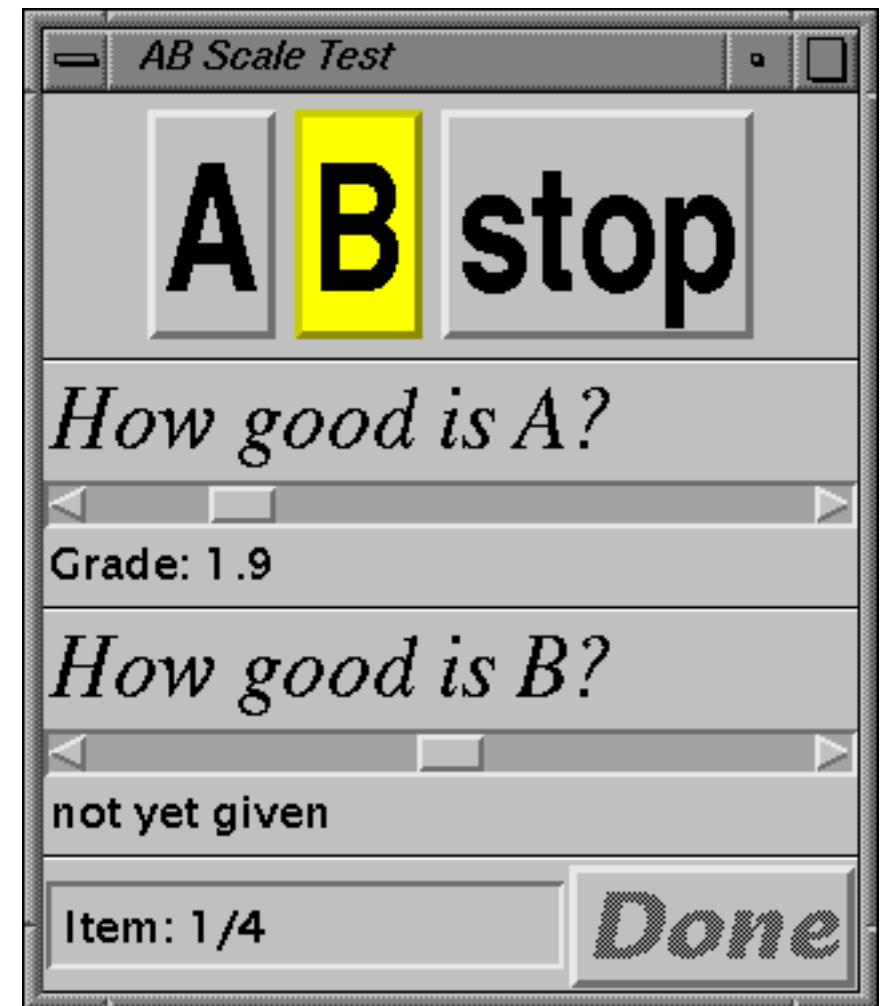
GuineaPig – Subject User Interfaces

- **Graphical answering panel for subjects** *(Java 1.1 AWT/Swing)*

- **Unlimited number of components can be added**

  - **Questions: many grades, multiple–choice, rank–order**

  - **Controls: sample play, button**

  - **Monitors: show test status, playing sample, time limit indicator**

- **Custom components fairly easy to add**

- **Multiple subjects simultaneously with remote terminals** *(not fully functional yet)*

- **Configuration with simple text files**
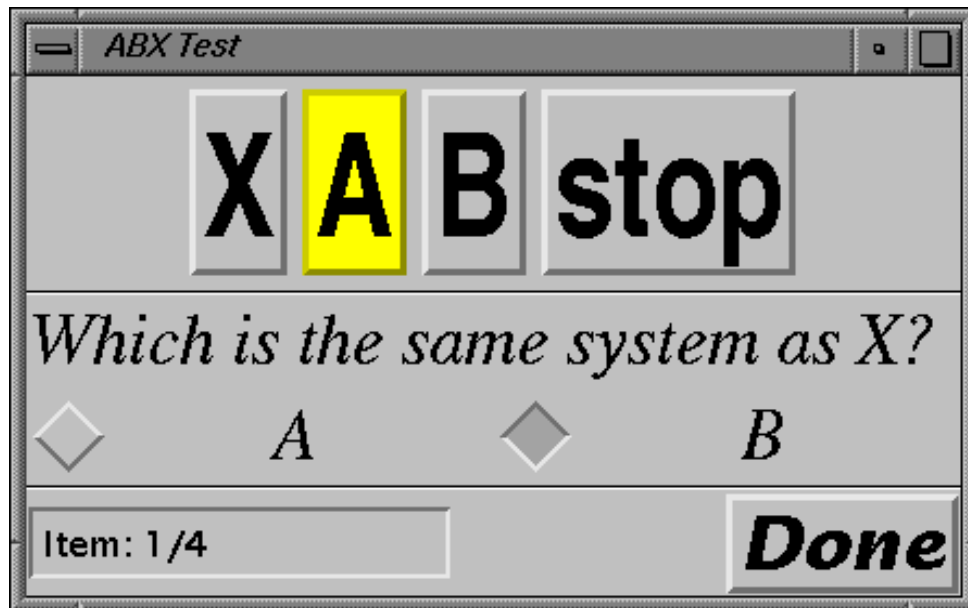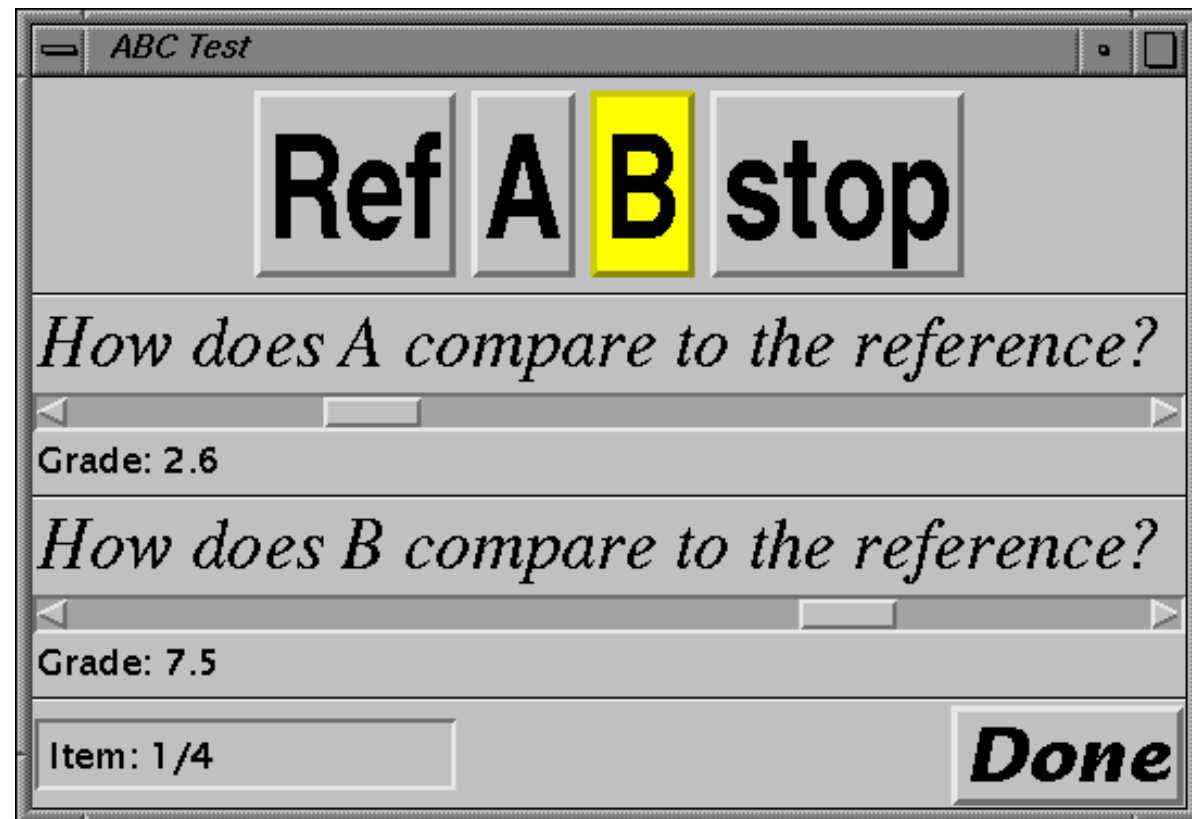
GuineaPig – Subject User Interfaces

A/B Scale test

A/B test

GuineaPig – Subject User Interfaces

Ref/A/B test

A/B/X test

GuineaPig – Subject User Interfaces

Rank order test

GuineaPig – Subject User Interfaces

Sample Play–controller / monitor

0–100 Point grading scale with 3 decimal places

ITU–R BS.1284, 5–point scale

ITU–T P.910, 9–point rating scale

ITU–T P.800, CMOS scale

Preference check boxes

Test progress status monitor ("Item: 4/20")

Answering time limit count–down indicator

GuineaPig – A generic subjective test system for multichannel audio
Presented by **Jussi Hynninen** at the 106th Audio Engineering Society Convention, 1999 May 8th, Munich, Germany

GuineaPig – Subject User Interfaces

Subject Panel config. example:

```
# Title of subject UI window
title: Meaningless test example

# List of question objects to be used.
questions:        question1,question2,question3,question4,q5

# List of control components to the used.
controls: play

# List of monitor components to the used.
monitors: play
```

```
# 'question1' is a grade using GradeBar.
question1.class:  GradeBar
question1.question:       Grade the clarity of sample A?
question1.minimum: 0.0
question1.maximum: 100.0
question1.decimals:       3
question1.questionfont:   Dialog-bold-18
```

**0–100 Point grading scale
with 3 decimal places**

```
# 'question2' is a five grade impairment scale
question2.class:  FiveGrade
question2.question:       What is the difference between sample B and C?
question2.questionfont:   Dialog-bold-18
```

**ITU–R BS.1284,
5–point scale**

... continues on next page ...

GuineaPig – Subject User Interfaces

... continued from previous page ...

```
# Example of emulating a TenGrade with a GradeBar
question3.class=GradeBar
question3.question=Rate the speech quality of sample D
question3.questionfont=Dialog-bold-18
question3.minimum=0.0
question3.maximum=10.0
question3.decimals=0
question3.choiceformat=1.0#Bad|3.0#Poor|4.0#Fair|6.0#Good|9.0#Excellent
```

ITU–T P.910, 9–point rating scale

```
# Example of emulating a TenGrade with a GradeBar
question4.class=GradeBar
question4.question=Rate the speech quality of sample D against A
question4.questionfont=Dialog-bold-18
question4.minimum=-3.0
question4.maximum=3.0
question4.decimals=0
question4.choiceformat=-3.0#Much worse|-2.0#Worse|-1.0#Slightly worse|0.0#About the\
same|1.0#Slightly better|2.0#Better|3.0#Much better
```

ITU–T P.800, CMOS scale

```
# 'q5' is a multichoice.
q5.class: CheckboxChoice
q5.question:     Which sample do you prefer?
q5.answers:      A,B,C,D
q5.questionfont: Dialog-18
```

Preference check boxes

```
# A controller to play samples
play.class:      PlayPanel
play.choices:    A,B,C,D
play.labels:     A,B,C,D
play.font:       SansSerif-bold-48
```

Sample Play–controller / monitor

```
testcontrol.font: Dialog-bolditalic-24
```

GuineaPig – Subject User Interfaces

GradeBar – Generic numeric scale

Grade the clarity of sample A?

◁   ▷

Grade: 41.561

Parameters:

- **Minimum / maximum of scale**
- **Number of decimals**
- **Show / don't show numeric grade to subject**
- **Adjectives associated with ranges of values**
- **Initial value, fixed / random**

GuineaPig – Subject User Interfaces

FiveGrade – ITU–R five–grade impairment scale

What is the difference between sample B and C?

Grade: 1.0 – Very annoying

- **1.0 to 5.0 scale**

- **Anchors (adjectives) derived from ITU–R five–grade impairment scale given in *Recommendation ITU–R BS.1284***

Impairment	Grade
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very annoying	1.0

GuineaPig – Subject User Interfaces

CheckBox – multiple choice

Which sample do you prefer?

◇   A   ◇   B   ◈   C   ◇   D

- **Subject selects one of the labels**

- **Any number of labels can be defined**

GuineaPig – Subject User Interfaces

RankOrder – ranking samples

Rank order the samples by spatial sound quality (1 = lowest quality, 7 = highest quality)

A	B	C	D	E	F	G
3.	6.	1.		4.		

- Any number of labels

- Allow / don't allow ties

- Allow /don't allow incomplete ranking (not all labels have been given a rank) as answer

GuineaPig – Test results processing

- **No analysis of test data is performed**

- **Exports data for analysis by other statistical analysis tools**

- **Results are exported as tabulated ASCII–text**

```
#session id: SES03
#session start time: Thu Feb 25 15:07:06 GMT+02:00 1999
#session end time: Thu Feb 25 15:08:37 GMT+02:00 1999
#session MCLL: 0.0dB
#ItemID  SubjID   SesID     Time/s   Switch   A        B        gB     gA
item1    hynde    SES03     15.7     7        pirr44   pirr32   5.6    1.2
item4    hynde    SES03     18.6     3        pirr11   pirr16   6.5    3.6
item2    hynde    SES03     TIMEOUT  8        pirr22   pirr32          7.9
item3    hynde    SES03     14.3     3        pirr8    pirr11   7.0    2.0
```

- **Format of exported results is configurable**

GuineaPig – Test results processing

Information that can be exported:

For each test item:

- Item ID
- Subject ID
- Session ID
- Item start time
- Item duration

- Number of sample switches

- Item parameters

- Item answers

Also session information as fields:

- Session start time

- Session end time

- Session's MCL level

GuineaPig – Test results processing

Output configuration options:

- which fields to print

- order of fields

- custom formatting of special data fields

- sub–fields: break one answer into several fields

- simple filtering of items based on *item ID*, *session ID*, and *subject ID*

GuineaPig – In conclusion...

- **flexible, scalable test platform for generic audio (visual) subjective testing**
- **completely software based, SGI / Java**
- **customizable / extensible**

Future development plans include:

- **better support for more channels, multiple concurrent independent subjects**
- **DRT tools**
- **real–time filtering, video support**

```
www.acoustics.hut.fi/~hynde/GuineaPig2/
```