Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than things you do expect) Plautus (ca 200(B.C.)

Project no: 027787

DIRAC

Detection and Identification of Rare Audio-visual Cues

Integrated Project
IST - Priority 2

DELIVERABLE NO: D4.2
Prototype of the VR Environment for Inconsistency Perception

Date of deliverable: 31.12.2006
Actual submission date: 26.01.2006

Start date of project: 01.01.2006 Duration: 60 months

Organization name of lead contractor for this deliverable: **IL-HUJI**

Revision [1]

Project co-funded by the European Commission within the Sixth Framework Program (2002-2006)		
Dissemination Level		
PU	Public	
PP	Restricted to other program participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	X
CO	Confidential, only for members of the consortium (including the Commission Services)	

Detection and Identification of Rare Audiovisual Cues

Insperata accident magis saepe quam quae speres.
(Things you do not expect happen more often than things you do expect)   Plautus (ca 200(B.C.)

D4.2 Prototype of the VR Environment for Inconsistency Perception

The Hebrew University of Jerusalem (HUJI)

Abstract:

We investigated the integration of audio-visual events in a detection task, and performed initial measurements of people's ability to detect incoherent (or rare) events. Specifically, we tested how categorical relationship between a sound and a target object will affect the perception of a sound as belonging to a target object. To this end we built a virtual city and planted objects emitting different sounds, where the subject's goal is to find incoherent events in the virtual environment. Our initial results show that incoherent sounds belonging to the same category as the target objects were more readily detected. However, individual detection rates indicate that some incoherencies constructed with different category sounds were very successfully detected, while other incoherencies were poorly detected. Our future work is aimed to clarify the circumstances under which this happens.

Table of Content

1. Introduction

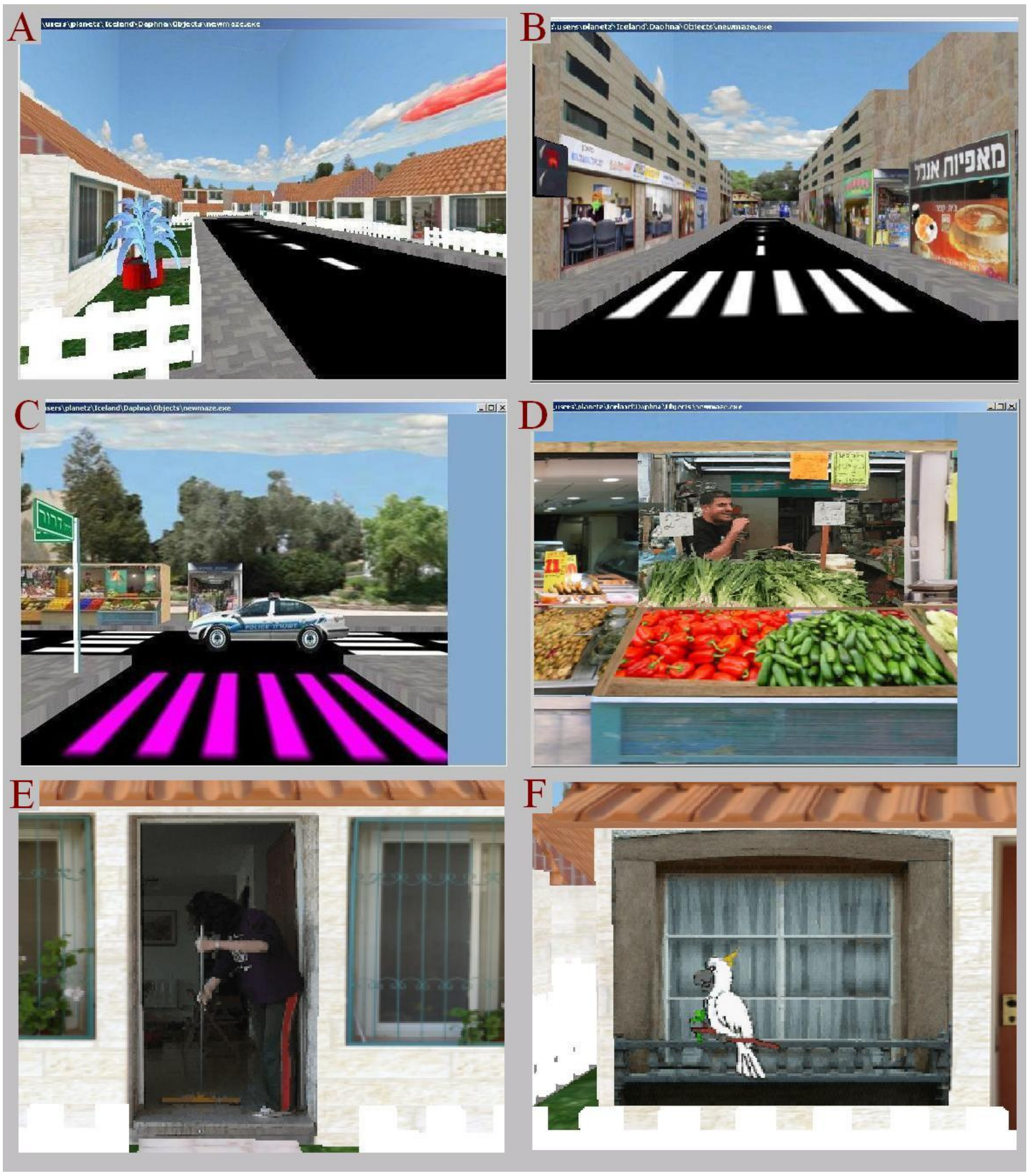

Figure 1. Incoherencies Detection Task Environment

A. A living neighborhood, a plant and a red cloud are seen. **B**. Shopping street. **C**. Police car going through an intersection with purple passing lane. **D**. A merchant making either lion (different category) or sneezing (same category) sound. **E**. A woman washing the floor making toilet flush sound (same category) or radio tuning sound (different category). **F**. A talking parrot sits in a window making rooster or sitar sound.

In this study we investigate the integration of audio-visual events in a detection task. Specifically we tested how categorical relationship between a sound and a target object will affect a perception of a sound as belonging to a target object. To test a perception of sound-object integration we built a virtual city and planted objects emitting different sounds, the subject's goal is to find incoherent events in the virtual environment. The detection of incoherent event will attest for perception of a "wrong" sound as belonging to a target object, otherwise the event won't be reported as incoherent.

For each object we created two versions of incoherent sound: (i) *same category* - an incoherent sound belongs to the same category as a target object, for example, street musicians playing on drums emitting trumpet sound (musical instrument -> musical instrument replacement), or a barking cat (animal->animal replacement); or (ii) *different category* - an incoherent sound belongs to the different category as a target object, as street musicians playing on drums emitting bubbling sound, or cat singing with woman's voice – different category (human->animal).

We hypothesize that categorically similar incoherent sounds will be detected more readily.

2. Methods

In this task, participants travel in a virtual city viewed using a Head Mounted Display (HMD) device. While the scene is intended to be as realistic as possible and to resemble our ordinary world, see Figure 1, along the way some 49 incoherencies are inserted, in which some object (or objects) are associated with inappropriate sound (18), color (14), or location (17). The participant is required to identify as many such incoherencies as possible. The participants were provided no information on three categories of possible incoherencies and didn't receive any examples.

Table 1. A full list of sound incoherencies used in the experiment

Object	Incoherent Sound	
	Same Category	Different Category
Dog	Cow	*Train horn*
Hammer	*Wood sanding machine*	Church bell
lawnmower	Fax	Bee
Baby banging a plastic can	Drums	*cuckoo clock*
Plane	*Bombing*	Horse galloping
Floor washing	*Toilet flushing*	Radio tuning whine
Parrot	Rooster	Sitar
Cat	*Dog*	Woman song
Fountain	*Rain-thunder*	Hairdryer
Ambulance	Ice-cream truck	*Marching feet*
Closing door	Breaking glass	*Car brakes*
Bus	Car brakes	*Elephant*
Drums	*Trumpet*	Bubbling sound
Merchant	Sneeze	*Lion*
Merchant	*Kissing sound*	Bird song
Child	Many children cheering	*Explosion*
Car	*Train*	Applause
Adults laughing	*Baby laugh*	Whistle

The sound incoherencies with same and different category sounds are listed in Table 1, in addition we included 15 normal sound events. For example, on the same category trial a dog mooed like a cow – the sound and object belong to the same category, and on the different category trial a dog made sounds of a train horn. 29 students volunteered to take part in the study and were randomly assigned to one of the two experimental conditions. Each participant experienced 9 incoherent sound events from the same category and 9 incoherent sound events from the different category. 14 subjects heard the sound indicated in red in Table 1 and 15 subjects heard the sounds shown in blue italics.

The Fountain incoherency was excluded from the final analysis because the sounds linked to it were not perceived as intended. The rain-thunder sound that was intended to represent the same category sound – a 'nature' sound, was perceived as the sound of a mechanical appliance.

3. Results

Incoherent sounds belonging to the same category as the target objects were more readily detected. The average detection rate of the same category sound incoherencies was 60.9% as opposed to 52.3% for the different category incoherent sounds. This 8.6% difference was highly significant, $p = 0.002954$. Figure 2 shows same and different category detection rates for each subject, and shows that the majority of the subjects had higher detection rates for same category sounds. Only 5 out of 29 participants (as indicated by dashed lines) had higher detection rates for sounds from a different category. No significant difference in overall performance was found between the two experimental conditions ('red' and 'blue' sounds in the Table 1).

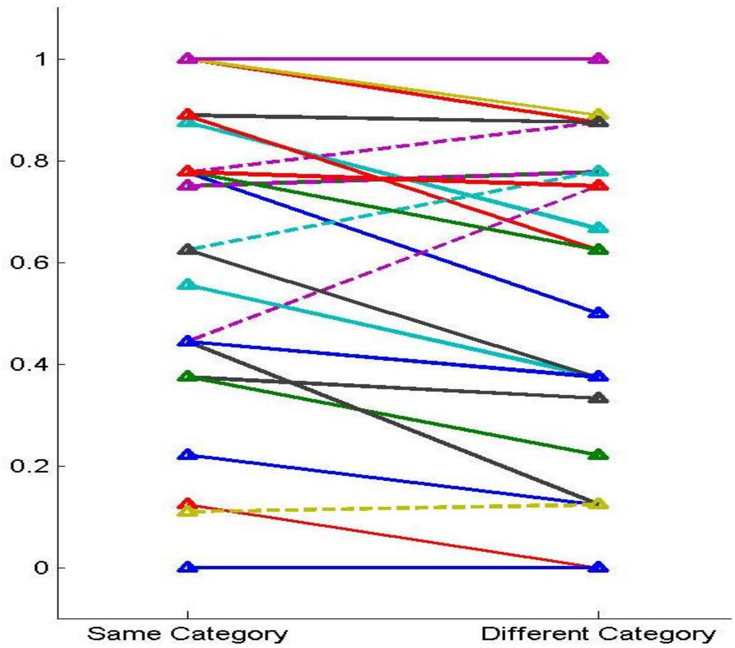

Figure 2. Sound Incoherencies Detection Rates of Same Category Sounds vs. Different Category Sounds in Normal Subjects

The detection rates for the same and different category sounds are shown for each subject. Solid lines indicate subjects that had better same category detection rates. Dashed lines indicate subjects that had better different category detection rates.

A comparison of detection rate of the incoherent sound from the same and different categories was also made for each incoherency. This also showed better detection of incoherencies when a sound and an object belonged to the same category (Figure 3). Specifically, 11 sound incoherencies were better detected when a sound belonged to the same category vs. 6 incoherencies that were better detected when a sound belonged to a different category.

The two most detectable incoherencies belonged to the same category: the ambulance playing the ice-cream truck jingle (93% detection) and a barking cat (89%). However, well detected incoherencies with a rate of 70-80% included 4 events from the same category (mooing dog, plastic can sounding like cymbals, talking merchant making a kissing noise, crowing parrot) and 5 from the different category, the latter included a bus trumpeting like an elephant, bubbling drums, a merchant roaring like a lion, a hammer pounding like bells and a plane galloping like a horse. The least successful incoherencies with detection rates below 40% were a door slamming with the sound of car brakes, a chirping merchant, a child making an explosion sound, and dog barking like a train – 4 events from the different category, and 3 events from the same category: a child sounding like many children cheering, a car accompanied by a train whistle, a hammer sounding like a wood sander.

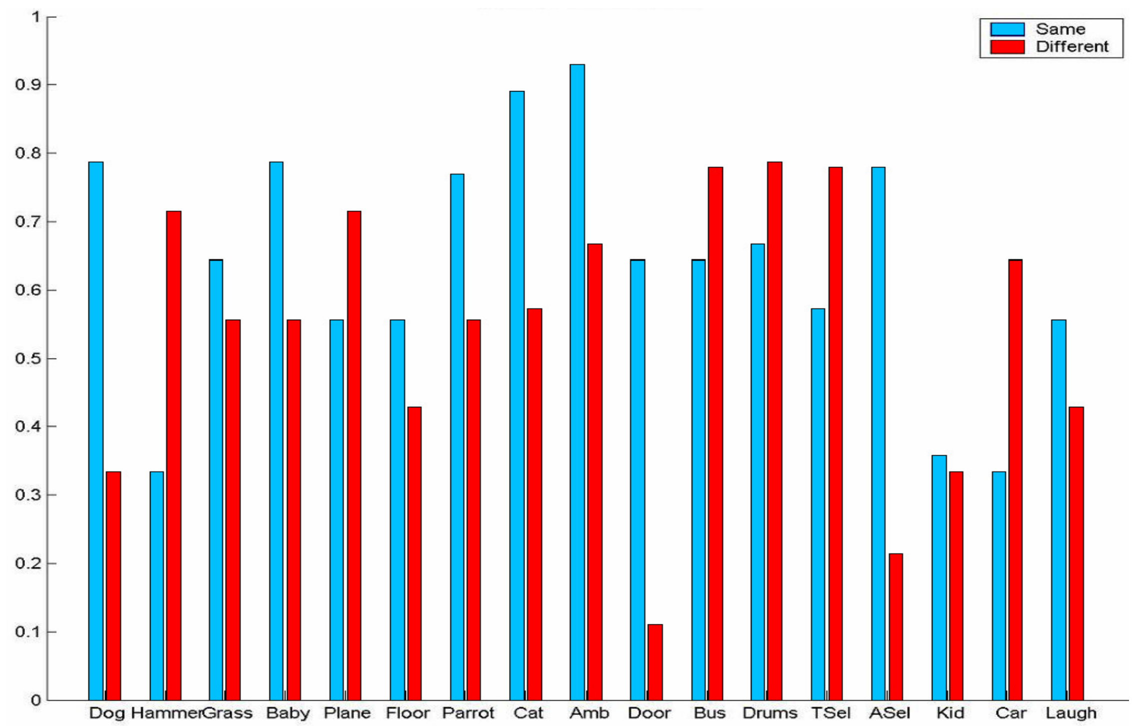

Figure 3. Sound Incoherencies Detection Rates of Same Category Sounds vs. Different Category Sounds for each Incoherency

The plot shows detection rates for sounds from the same category (blue) and from the different category (red) for each of the 18 incoherent objects, see Table 19 for specific sounds.

4. Discussion

Incoherent sounds belonging to the same category as the target objects were more readily detected; this was exhibited both in individual subjects' detection rates and detection rates for each incoherency. However, individual detection rates indicate that some incoherencies constructed with different category sounds were very successfully detected, such as the merchant roaring like a lion or the 'elephant' bus, and some same category incoherencies were poorly detected, such as a car making a train sound.

The division into categories is not well defined, and other criteria may better predict which sounds will be better detected, for example, a sound relating to the same action. The baby banging the plastic can accompanied by a sound of cymbals is in fact best described as the same action rather than the same category. For a few incoherencies the different sound was very successful in being perceived as emitted by an object, such as hammer with bells (vs. sander) sound, and the merchant with lion (vs. sneezing) sound.

Another factor that may affect the perception of a sound source is whether the sound is likely to be heard against background noises on the city streets or not. The biggest difference between same and different category detection rates was seen in the dog with cow or train sounds, the closing door accompanied by breaking glass or squealing car brakes, and the merchant making kissing sounds or bird chirps. In all these cases the different category sound could be attributed to ambient city noises. On the other hand, a lawnmower sounding like a fax or buzzing like a bee had little effect on detection rate. Applause accompanying a passing car or elephant trumpeting associated with a bus are unlikely to be heard on the streets and yielded higher detection rates than same category sounds. The incoherent sounds that are rarely expected to be heard on the streets were more readily detected, 65% on average (SD=13), as compared to common street sounds – 54% on average (SD=23), this difference is close to significance (F=3.14, p=0.09). Categorical resemblance, similarity of action, sound frequency and expectancy in every day life environments - these are probably only some of the factors that affect sound perception and linking to a source object. We had no controls for these situations.