

Multimodal Perception: Part 2
Professor Benjamin Weiss
Quality and Usability Lab
Technische Universität Berlin
Effects of Disgruent Signals

(Refer Slide Time: 00:16)

Multimodal Perception

Outline:

- Processing multiple signals
- Multimodal Dual-Tasks
- Effects of discongruent multimodal signals
- Relevance



If we want to find out how multimodal perception is done in humans, so how different signals are processed cognitively, it is a nice way to actually synthesize signals which have not the same information.

(Refer Slide Time: 00:35)

Discrepancies in multimodal integration.

- General observation: If signals – expected to be from the same origin – are not congruent → humans adapt them as long as there is no striking evidence they are not from the same origin!
- cross-modal perception: ask for one modality
- multimodal integration: ask multimodal
- assumption of unity!



Imagine for example you are sitting in a train at a train station. And on the next platform there is also a train and you can see from the corner of your eyes the train moving. You expect

really that your own train is moving and may be for some milliseconds you actually feel the acceleration, until you realize your train is still standing but the other train has started.

So here you can also see that if we have different information, this case the balance sensor and visual sensory information, they are not congruent. Then we use these information and process them anyway to actually form something often valid or not valid picture of how the world works. We have several of these kind of incongruent signals that we are dealing with.

(Refer Slide Time: 01:29)

Discrepancies in multimodal integration.

Visual effect on sound intensity (Fassi & Zwicker, 2007)

- Description: Visual context affects loudness perception
 - Example: "Red cars" sound louder than green ones
 - Result: One modality affects the perception of another
- cross-modal perception



So for example there is an interesting effect that if we provide different cars and, in different colors and these are also producing some noise, the typical car noise. If we ask people how loud these cars are, we see an influence of the car color on the loudness perception.

(Refer Slide Time: 02:29)

Discrepancies in multimodal integration.

- Visual-kinesesthetic (haptic) location (Helmholtz, 1866)
- Description: Visual and kinesesthetic information are not congruent
 - kinesesthetic (haptic) information is adapted
 - Examples: prepared glasses when pointing on a target (works also for the shape of objects)
 - Result:
 - This effect exemplifies the plasticity of the brain to adapt to changes quite rapidly, the "after effect", when removing the glasses prove the adaption, kinesesthetic (haptic) information of the hands are not just overridden.
 - Only works with self moved hands/arms
 - recalibration of sensory integration by movements
 - firstly interpret as "modality dominance" of vision
 - multimodal integration



This case, for example red cars are on average perceived little bit louder than other colors. So this is an example of a crossmodal perception. This means we are asking for one some modality in this case, the auditory information how loud is it but we have an effect of another modality, the visual information color affects the loudness perception.

One of the older examples of these discongruent information is actually reported by Helmholtz. So the idea is that you ask people to point to a certain object which sounds reasonably enough. You see object and you point there.

But if you put on some glasses or distort the vision by a kind of lens, you can actually dislocate the object on the visual place, on the visual location. What happens here is that if we have to do that, the people adapt their pointing gesture to the visual information which are actually wrong or distorted. And they do this quite quickly.

This actually the origin of the so-called theory that vision dominates all senses which we shall later that it is, that it is not true. The most important aspect of this whole paradigm is that when we take off the glasses or lenses again, it takes some time until we adjust our kinesthetics, our knowledge about our body and posture so that our pointing gesture actually is correct again.

For certain amount of time, our pointing gestures after removing these glasses will be wrong. It will still be adapted to the initially wrong information from the glasses. So this only works if we actually move our arms actively, so we need this feedback from our body.

And this is an example of what I call multimodal integration because here we do not ask specifically for one modality or the other but we ask for the whole object. And ask for the people to integrate the information that they have.

So it is only a little bit different in the task paradigm and the instruction that the users or the participants of this experiment are given to.

(Refer Slide Time: 04:50)

Discrepancies in multimodal integration.

- Audio-visual effect on counting (Shams et al. 2000)
- Description: Counting of short events with auditory and visual information mismatch
 - Example:
 - Three visual stimuli in a row (flashes)
 - How many flashes?
 - More auditory stimuli → More flashes perceived
 - <http://www.cns.uci.edu/~kmtm/audvisual/Rabbit/index.html>
 - <http://shamslab.psych.ucda.edu/demos/>
 - Result: Audio is dominant
 - cross-modal perception



There is also an audiovisual effect on counting. The whole idea behind this example is that we see visually appearing certain kinds of objects, in this case rabbits or white flashes. We also have certain inputs, sounds which are occurring sometimes in the same time, same timepoint as the flashes.

But the number, location of the sound stimuli and the visual stimuli are not congruent, the question is which of these modalities is actually dominating the impression of the people?

I will give you a short example on the website.

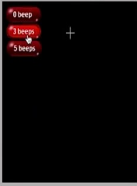
(Refer Slide Time: 05:44)

Sound-induced Visual "Rabbit": Demos

Kamitani, Y. & Shimojo, S.(2001) Sound-induced visual "rabbit". *Journal of Vision* (in press).


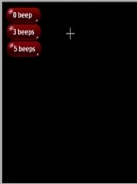
Three bars accompanied by a different number of beeps

Fixate to the cross, and press one of the buttons. You may see a different number of bars with a different number of beeps (especially with '3 beeps'), whereas the visual stimuli are physically identical.
(This demo may not work on slow computers. If the bar and beep are not synchronized with '3 beeps', you cannot observe the effect, sorry...)



Three rabbits accompanied by a different number of beeps

Why do we call it sound-induced visual "rabbit"? See Gelbard and Sernick (1972). But it can be easily made more like rabbits as below :)



0:05:43.0 Demo start

0:06:00.8 Demo end

(Refer Slide Time: 06:04)

Discrepancies in multimodal integration.

- Audio-visual effect on counting** (Shams et al. 2000)
- Description: Counting of short events with auditory and visual information mismatch
 - Example:
 - Three visual stimuli in a row (flashes)
 - How many flashes?
 - More auditory stimuli → More flashes perceived
 - <http://www.cns.afj.jp/~kmi/audiovisualRabbit/index.html>
 - <http://shamslab.psych.uct.ac.za/demo/>
 - Result: Audio is dominant
 - cross-modal perception



I do not know what you have perceived but the majority of the people actually was influenced by the auditory information, so these were dominating the visual ones.

And this is one example again of crossmodal reception at least if you ask for the visual number of occurrences of rabbits or flashes. Of course you can also ask about how many objects are occurring here then it would be multimodal integration.

(Refer Slide Time: 06:35)

Discrepancies in multimodal integration.

- Audio-visual effect on object identification** (Sekuler et al., 1997)
- Description: Auditory context effects visual perception
 - Example: Identical objects pass through each other, but bounce with sound
 - Result: One modality effects the perception of another
- cross-modal perception


• http://www.michaelbach.de/ot/mot_bounce/bounce.swf
• http://www.michaelbach.de/ot/mot_bounce/index.html



The next effect is on audiovisual object identification. And again I have short example. You will see two balls or two circles running towards each other. And either they are crossing or they are bouncing off dependent on the auditory information. Have a look yourself and find out what you are perceiving here.

(Refer Slide Time: 07:06)

next > **Motion-Bounce Illusion** < prev

from Michael's Visual Phenomena & Optical Illusions

Reset

Soundcheck

1

2

-2 < > other

5 Speed

7

Build with...

©2002-17 Michael

Warning: this is a subtle effect, and you have to calibrate your browser. [Doesn't work?](#)

In order to appreciate this, please proceed in the following order:

1. Test your sound output by pressing the button 'Soundcheck' (or hit key 'V'). You should hear something of a 'clack' sound. If not, up your sound volume or go to another illusion.
2. Calibrate the sound delay of your browser* as follows: press the button labeled '1' (or hit key '1'). The blue balls will begin to move smoothly. At the very moment they meet, the 'clack' sound should occur. Use the long vertical slider to adjust the timing. If the sound occurs too late (or too early).
3. Try out the actual phenomenon like so:
4. Calibrate the effects after pressing buttons '1' or '2' and lib by determining the paths of the blue balls. Are they crossing like so > <? or do they bounce off each other > <?

Changes in the sound when there is no accompanying sound, the crossed path is seen, with the sound (> <) is more likely to be perceived.

0:07:07.1 Demo start

0:07:21.4 Demo End

(Refer Slide Time: 07:23)

Discrepancies in multimodal integration.

Audio-visual effect on object identification (Sekuler et al., 1997)

- Description: Auditory context effects visual perception
 - Example: Identical objects pass through each other, but bounce with sound
 - Result: One modality effects the perception of another
- cross-modal perception



- http://www.michaelbach.de/otmot_bounce/bounce.swf
- http://www.michaelbach.de/otmot_bounce/index.html



So the result is of course, that once there is bouncing sound we perceive this as two objects bouncing off each other and not going through each other. So this can be nicely used if you want to synthesize certain behavior objects by the help of auditory information.

And again as we are asking for visual information of the auditory, of the object, the identification, which object is which one, are they bouncing off or are they going through each other, this is also an example of crossmodal perception.

(Refer Slide Time: 08:06)

Discrepancies in multimodal integration.

Audio-visual discrepancy in emotion (de Gelder et al., 1995)

- Description: Emotion in audio effects visual emotion identification (and vice versa)
 - Examples: Sadness versus Happiness with neutral semantics
- Result:
- Identification lowered with mismatched information
- Not effected by attention or explanation
- cross-modal perception



The next effect is on audiovisual emotion or multimodal emotion recognition. I do not have an example here but if we present pictures or video clips of faces of people with certain emotion and we also present vocal stimuli of people where you can also perceive

(Refer Slide Time: 08:32)

Discrepancies in multimodal integration.

Audio-visual discrepancy in emotion (de Gelder et al., 1995)

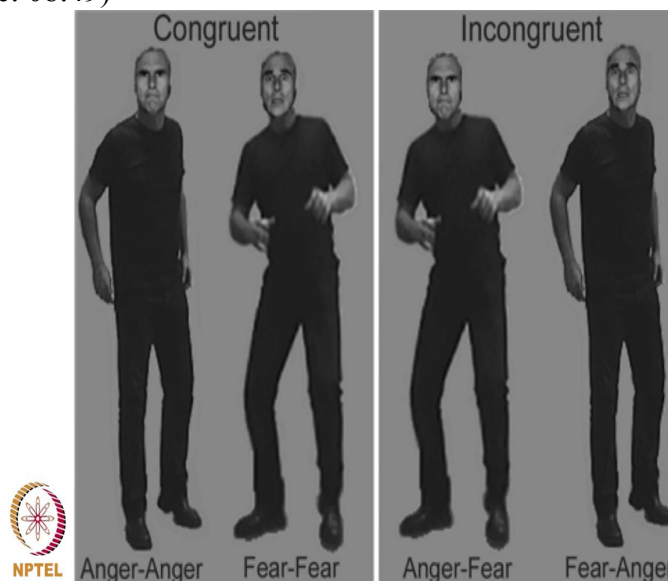
- Description: Emotion in audio effects visual emotion identification (and vice versa)
- Examples: Sadness versus Happiness with neutral semantics
- Result:
 - Identification lowered with mismatched information
 - Not effected by attention or explanation
 - cross-modal perception



the emotion that the person has, you will have a certain crossmodal effect.

So if you ask for example, of the emotion by pictures, there will an effect of the auditory information and vice versa.

(Refer Slide Time: 08:49)



Here you a certain example with not auditory information but with visual information on face and gesture or posture of these people.

On the left side you can see congruent information, so anger posture and anger face and fear posture and fear face. On the right side you see incongruent information.

That is not truly multimodal as both information are transported on visual domain but illustrates nicely that if we now ask for the information there will be an effect of posture on the facial expression or vice versa.

(Refer Slide Time: 09:33)

Discrepancies in multimodal integration.

Audio-visual location ("ventriloquist illusion", e.g. Harris, 1965)

- Description: Auditory and visual information are not congruent → visual information often dominates, auditory information is then adapted
 - Examples: ventriloquist, "obvious" loudspeakers
 - Result:
 - Effect is of course dependent on the spatial difference (assumption of unity)
 - Effect dependent on intensity of auditory and visual stimulus (e.g. Radeau, 1985)
- multimodal integration



The next example is the most famous one. It is about audiovisual location and the typical example is the so-called ventriloquist illusion. This means we have an actor on stage who is talking but he or she has a puppet nearby and we perceive this in the audience as the puppet is speaking.

So there is a discrepancy in the location because the auditory information is coming from the person, the actor but the visual information is the moving lips and the head of the puppet.

So here we have natural environments typically the dominance of the visual information. But we know that this can be blurred leading to an interesting result if we do this in the so called, in the studio or laboratory; but just finish with this audiovisual location illusion.

So this would be a nice example of multimodal integration because these two information are processed together to form the whole origin, the location of this person who is speaking.

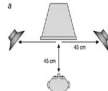
(Refer Slide Time: 10:51)

Discrepancies in multimodal integration.

Audio-visual location

- Alais & Burr, 2004
- Screen with 2 speakers (stereo); interaural time difference
- Sound: 1.5ms stimulus
- Vision: gaussians with various constants (blobs)
- Ventiloquist effect not for differences over about 40°
- Accuracy of senses seem to play a role (light "blobs" & sound "clicks")
 - If the visual information is blurred much (about 10°), auditory information may dominate
 - If the visual information is blurred to a precision similar to acoustic information, both information are averaged → optimal integration

→ multimodal integration



In the laboratory you can try to reduce the reliability of the visual information on the screen. So for example we have the setup with two loudspeakers so we can actually try to simulate the different sound sources by having a certain delay of the right or the left channel.

And we have the position of so-called blob or blurred light on the screen. And we blur this information which is not as sharp as the moving mouth of the puppet by some Gaussian distributions, by introducing some noise. So there are no clear boundaries of this white blob occurring somewhere.

And if we test this we find interestingly that the more we blur, the more we reduce the visual information reliability of that, the more important the auditory information gets, until of course if the information are too far away or too incongruent there is no assumption of unity any more.

So the main result of this laboratory experiments shows us that there is no dominance of visual information but rather our processing can be presented or can be called a kind of optimal integration. So we actually use from experience how reliable the source of information is that we have.

So given a certain situation, giving our knowledge about the reliability and accuracy of our senses and the information that we perceive, we weight the signals, we weight the information from the different sources or senses and form kind of optimal integration.

This can actually result in having a result final multimodal perception in this case of the location which is neither on the same, the same or identical with the auditory nor with the visual information.

So for example if we have a properly blurred visual location and an auditory location we might perceive the actual object being in the middle.

(Refer Slide Time: 13:22)

Discrepancies in multimodal integration.

Audio-visual location/quality ("McGurk effect", McGurk und MacDonald, 1976)

- Description:
 - Auditive: /baba/
 - Visual: /gaga/ /dada/
 - Audio-visual: /dada/ /dada/
- Examples: ECAs and designed videos
- Result:
 - fusion, multimodal integration
 - New articulatory "location"
 - Voicing not affected (/d/, not /t/)
 - But: vision: /baba/, auditory: /dada/ → /bdabda/ (combination, not fusion)

Check out the video without comments from Prof. Kuhl



This kind of optimal integration is also the result of the next example.

This is called the McGurk effect or the audiovisual location or audiovisual quality. This is about sound perception. There is again a nice video. A professor also explained what you should hear, what you should do. So please check it out. It is on the website and come back once you have finished that.

I do not know if this experiment worked for you but for most people it does and it is an automatic process. So usually it works although you know how it works, it will be repeated, can be repeated and will work again.

So the whole idea behind this is that our auditory perception perceives a /ba in the vocal tract, in the mouth which is not very accurate and the visual information is a /ga which is far back in the mouth.

So in order to combine this information to one meaningful location or one meaningful articulation on sound the only thing that we can do is find the sound between these two discongruent information.

And this is usually /da or /ba as this relies on the accuracy of the different information. You might imagine what happens if we now reverse the kind of information.

So if we would provide visual information of /ba and auditory information of /ga, this case it would not work. Because the visual information of /ba is really strong, very salient and reliable. So the result might be something like /da/da.

So as I said, again this is a nice example that there might be some modalities which dominate the perception; for spatial information, usually the vision but in the real process behind our multimodal integration is so-called some kind of optimal integration that means each modality is weighted to its situation and the reliability and expectation of the quality of this information.

So in the end, multimodal perception might be some kind of information that is not available in either of the two channels and sensory information that we have but it would be the most meaningful one, always assuming the unity of same sound and vision source of the signals.

(Refer Slide Time: 16:41)

Discrepancies in multimodal integration: Explanation Theories.

- **Visual dominance**
 - In many early studies: vision dominates other modalities (audio, kinesthetic)
 - But: mostly spatial tasks, where vision is exact and reliable
- **Modality appropriateness**
 - Situational appropriateness determines dominant modality
 - E.g. audio for timing, visual for spatial tasks
- **Other integration approaches**
 - Optimal integration → a weighting dependent on the estimated accuracy of information; might result in an averaged result
 - Bayesian modeling, maximum likelihood etc.



So as I have already mentioned, the first kind of theories were that vision dominates all other senses. But with the temporal domain and the counting there was also the result that for some experiments and some observations auditory information are dominating.

So as a second kind of explanation theory there was the so-called Modality Appropriateness Theory just claiming that the one modality dominates the other or others which in this kind of situation is most appropriate.

But now we have seen two examples, the McGurk effect and also the ventriloquist effect which shows that the resulting percept might be different from all the kind of information that we are presented on the different senses. And therefore the coherent explanation theory is the so-called Optimal Integration.

(Refer Slide Time: 17:53)

Discrepancies in multimodal integration: Summary.

- There is the assumption of unity
- Crossmodal effects: perception of one modality is affected by another
- Multimodal effects: multimodal (holistic) perception is not always the sum of single signals
- Several effects give evidence for:
 - Automatic process of cross-modal perception
 - Not a conscious process
 - Information of different sensory inputs are evaluated together to obtain a valid, robust impression of the world
 - Not more costly than uni-modal processing (Kinsbourne, 2003) (in contrast to some dual-task paradigms)
- Which is the dominant modality in the case of conflicting information depends on
 - Situated "quality of the modality" (reliability, "prior knowledge"?)
 - "attention towards this modality"



As a summary, in order that these kind of effects work, we need the assumption of unity. So the information should not be too discongruent from each other. I showed two kinds of effects.

These are actually only related to the kind of tasks the people have in the experiment, either there is crossmodal reception when we ask for one modality and we see that information from other modalities affect the results from the participants although they should not.

And also there is multimodal integration when we just ask for the overall multimodal perception. All these kind of effects give us reason to believe that the processing, the cognitive processing of the multimodal input is mostly automatic and not conscious.

And it deals with the meaningful convergence, or the meaningful processing of all the information we have in order to deal with the world. So expectations highly affect the way we process our information and the results especially concerning the patterns that we see and that we perceive.

So there are lot of other mostly visual nice effects showing that the expectation governs and affects what we actually perceive in end.

Another result is that this kind of perception is not more costly than unimodal processing. And we know that reliability and accuracy, but also our current attention is relevant for the outcome of these kind of multimodal integrations.