

Multimodal Action
Professor Benjamin Weiss
Quality and Usability Lab
Technische Universität Berlin
Gestures to Space

(Refer Slide Time: 00:15)

Multimodal Action

- Outline:**
- Introduction: Verbal and nonverbal information
 - Gestures
 - Posture
 - Gaze
 - Voice
 - Space
 - Turn taking
 - Emotion
 - Social relationship
 - Summary



Now that I have introduced what non-verbal information are and what verbal information are, and why these are important for human computer and also human interaction, I want to start presenting the most important kinds of non-verbal signals and non-verbal information for this week.

(Refer Slide Time: 00:33)

Gestures

- Phases of phrase: preparation → stroke → retraction (optional), (cf. Kendon, 2004)
- Gestural phrases may be connected to gestural units (starting and ending with resting phrases)
- Gestures are synchronized with speech
 - Gestural phrase, intonation phrase (without retraction)
 - Stroke, onset of stressed syllable
 - Gestural unit, whole sentences (with retraction)
- Functional classification of gestures
 1. Emblems (cultural dependent explicit gestures (e.g., nodding, thumb up))
 2. Affect display (completion of language)
 3. Adaptors (playing with something, scratching the head)
 4. Speech-related / Illustrators



I will start with gestures which are most likely the most important non-verbal information that we can use in human computer interaction. But typically gestures are currently used with

touch screens, for example or with cameras to operate the system by predefined commands. This means the user has to learn which gestures to use for operating system or for certain functionality.

But of course if I am producing speech, engaging in a conversation I naturally use gestures without a certain predefined function for that. That means all these, there is a high variability in the gestures that we have. But still we have actually 3 kinds of phases of a gesture.

We are beginning with the preparation. This just means that the necessary part of our body, in this case mostly hands and arms are moved in a starting position. In order then in the second phase and the major phase to produce the real stroke and this is the archetypical or prototypical movement of the gesture that is iconic or relevant for the function, meaning or the, at least the characteristic of the gesture.

Then we have the third optional phase. This means we can retract our hands and arms again to kind of resting position somewhere else, but of course this is optional because we can have a sequence of these gestures to some form a gestural unit.

In speech production we assume that the gestures and the verbal information, the verbal gestures of our articulators are actually based on the same fundamental, main fundamental representation which is rather abstract.

This means when you produce a sentence that, that our gestures are not using any kind of additional cognitive resources but it is natural to do that in accordance and even in high temporal synchrony with our utterances what we are speaking. So there are certain loose relationships that we can define between gestures and speech.

For example, the gestural face is usually or loosely correlated with the so-called intonation phrase which is for in other terms, something like a spurt. So this means it is kind of a sentence or utterance that I can produce without having a certain break or without a silence or without the need for air and breathing.

The stroke itself is often highly correlated and synchronized with the onset of a stress syllable or the sentence stress. And then we have a loose correlation again between a gestural unit and the whole sentence which might include several intonation phrases or several spurts.

I also want to present a small classification of gestures and I will go into details on the next slide. So we have emblems. Emblems are actually rather independent of speech, or at least they can be independent.

And this means we have highly learned cultural dependent signals or signs. For example, for thumbs up for that was great, or for hitch hiking or nodding for saying yes or indicating yes.

These are however learnt, they can be used in isolation but also with speech and they are cultural-dependent. So in another culture, another country they might have a totally different meaning. These are called emblems.

Then we have affect display, something like facial expressions but also my posture and of course something like uncertainty, all these gestures are related to emotion and affect. Then we have so-called adaptors and these are defined by having nothing to do with the semantic meaning of the sentence that I am actually uttering.

For example, if I am playing around

(Refer Slide Time: 05:09)

Gestures

- Phases of phrase: preparation → stroke → retraction (optional), (cf. Kendon, 2004)
- Gestural phrases may be connected to gestural units (starting and ending with resting phrases)
- Gestures are synchronized with speech
 - Gestural phrase , intonation phrase (without retraction)
 - Stroke , onset of stressed syllable
 - Gestural unit , whole sentences (with retraction)
- Functional classification of gestures
 1. Emblems (cultural dependent explicit gestures (e.g., nodding, thumb up))
 2. Affect display (completion of language)
 3. Adaptors (playing with something, scratching the head)
 4. Speech-related / illustrators



with something that is making me discomfort, for example,

(Refer Slide Time: 05:18)

Gestures

- Phases of phrase: preparation → stroke → retraction (optional), (cf. Kendon, 2004)
- Gestural phrases may be connected to gestural units (starting and ending with resting phrases)
- Gestures are synchronized with speech
 - Gestural phrase , intonation phrase (without retraction)
 - Stroke , onset of stressed syllable
 - Gestural unit , whole sentences (with retraction)
- Functional classification of gestures
 1. Emblems (cultural dependent explicit gestures (e.g., nodding, thumb up))
 2. Affect display (completion of language)
 3. Adaptors (playing with something, scratching the head)
 4. Speech-related / Illustrators



with my hair, of course I can, it might be similar to telling somebody intentionally that I am thinking about it but could be just that I have an itch here.

And if I have an itch here and that is nothing to do with my verbal production then it is called an adaptor. Another example would be playing around

(Refer Slide Time: 05:37)

Gestures

- Phases of phrase: preparation → stroke → retraction (optional), (cf. Kendon, 2004)
- Gestural phrases may be connected to gestural units (starting and ending with resting phrases)
- Gestures are synchronized with speech
 - Gestural phrase , intonation phrase (without retraction)
 - Stroke , onset of stressed syllable
 - Gestural unit , whole sentences (with retraction)
- Functional classification of gestures
 1. Emblems (cultural dependent explicit gestures (e.g., nodding, thumb up))
 2. Affect display (completion of language)
 3. Adaptors (playing with something, scratching the head)
 4. Speech-related / Illustrators



with the, with the pen here.

The most important or the most interesting class of gestures are the speech-related one also called illustrators. So let us go into details here.

(Refer Slide Time: 05:52)

Gestures

- Referent-related Gestures (spatial information)



We have certain referent related gestures, for example spatial information.

So if I am pointing somewhere this would be such kind of a speech-related and speech-companion gesture. That means I can easily use it within my or this modification or to

(Refer Slide Time: 06:15)

Gestures

- Referent-related Gestures (spatial information)



support my utterances here.

(Refer Slide Time: 06:18)

Gestures

- Referent-related Gestures (spatial information)
- Relationship (shrug to signal uncertainty, showing the palm to plead)



FIGURE 7-9
Palm gestures.



Taken from Knapp & Hall, 2010

I can also signal relationship. Here is an example of kind of uncertainty in what I am actually delivering to my interaction partner.

(Refer Slide Time: 06:31)

Gestures

- Referent-related Gestures (spatial information)
- Relationship (shrug to signal uncertainty, showing the palm to plead)
- Punctuation (emphasize: point up & organize: 1,2,3.)



FIGURE 7-11
Punctuation gestures.

Taken from Knapp & Hall, 2010



And also, there is of course punctuations. And these are nice because they are highly temporally synchronized with my speech.

For example, if I am now telling you that I have 4 of these kind of gestures, I have first, referent signals, second, punctuation, third and so on. These are temporally synchronized with my speech.

(Refer Slide Time: 06:54)

Gestures

- Referent-related Gestures (spatial information)
- Relationship (shrug to signal uncertainty, showing the palm to plead)
- Punctuation (emphasize: point up & organize: 1,2,3.)
- Interaction

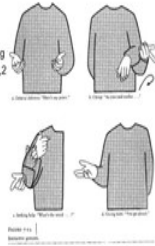


And then there is interaction of these kind of signals, of these kind of gestures which are actually used for regulating who is going to talk and who is going to listen.

(Refer Slide Time: 07:07)

Gestures

- Referent-related Gestures (spatial information)
- Relationship (shrug to signal uncertainty, showing
- Punctuation (emphasize: point up & organize: 1,2
- Interaction



Taken from Knapp & Hall, 2010



Here is an example of such an interactive gesture.

(Refer Slide Time: 07:16)

Gestures

- Semiotic classification of gestures (cf. McNeill, 2000)
- Deictic (pointing)
- Iconic (concrete idea: attributes "so big", relations "right beneath it")
- Metaphoric (abstract idea: "talk was so long" - "rolling" movement)
- Beats (rhythmic movement)



There is also another approach, a more semiotic approach by McNeill

And he is classifying gestures according to their semiotics and that means for example, we have again pointing gestures, deictic gestures. We have also iconic gestures and I have already given you an example with the big fish; that may be like this or like this.

These iconic gestures are rather concrete. A little bit more abstract are metaphoric gestures. For example referring to something that was said earlier or expressing that time goes by, these would be metaphoric gestures. And then again we have beats which resemble the punctuations.

(Refer Slide Time: 08:08)

Gestures

- Social signals are cultural dependent
- E.g. Study by UXfolios.com
- 300 apts. in 18 countries
- Limited set of universally usable gesture available
- Channel up/down
- Navigate in menu
- Volume up/down
- Problematic for advanced features
- Share program

We especially observed some other creative approaches to advanced TV functions.

Share current program, Argentina



Share current program, China



Share current program, Korea



Electronic program guide, Belgium, UK



Contact info for current program, Belgium



Contact info for current program, Spain









So gestures are culturally dependent.


(Refer Slide Time: 08:16)

Gestures

- Social signals are cultural dependent
- E.g.: Study by UXfellows.com
 - 360 subj. in 18 countries
 - Limited set of universally usable gesture available
 - Channel up/down
 - Navigate in menu
 - Volume up/down
 - Problematic for advanced features
 - Share program

We repeatedly observed some rather creative approaches to advanced TV functions:

Share current program, Argentina	
Share current program, China	
Share current program, Korea	
Electronic program guide, Belgium, UK Context info for current program, UK	
Context info for current program, Belgium	
Context info for current program, Spain	



This means actually a big obstacle for human computer interaction. And here is a small figure summarizing inter-cultural study on how to use gestures for Smart TV and video center.

I do not want to go into detail. You can easily go on the website of these UXfellows and look at the study yourself. But the main result was that the more concrete the functions are, that the users are using to, to map gestures to the functions.

For example, to navigate to the next channel, to raise the volume of your media center and so on. So these are all really predefined pre-learned command gestures and the only study goal was to find out which one would be the best fitting or most intuitive ones. Here we find that they are, this is easy to find a common gesture for these really concrete easy functions.

But the more complex the functions get, for example navigating in a certain menu or complex task like recording and so on then people will diverge a lot and is special between the different national cultures. So it is actually an issue to come up with a nice gestural interface for operating devices.

(Refer Slide Time: 09:53)

Multimodal Action

- Outline:
- Introduction: Verbal and nonverbal information
 - Gestures
 - Posture
 - Gaze
 - Voice
 - Space
 - Turn taking
 - Emotion
 - Social relationship
 - Summary



After gestures I want to talk about posture.

(Refer Slide Time: 09:58)

Posture

- Coordination
- Gestures, postures, facial expressions and
 - Self-synchrony
- Interaction synchrony (alignmer)
- Interaction coordination
- Back-channeling
 - Turn-taking



So as with gestures the whole body orientation, the posture is also highly coordinated with the speech. For example they are self-synchrony.

(Refer Slide Time: 10:13)

Posture

Coordination

- Gestures, postures, facial expressions and
- Self-synchrony

• Interaction synchrony (aligner)

• Interaction coordination

- Back-channeling
- Turn-taking

FIGURE 7-14 | FIGURE 7-15
Some postural-kinetic markers | Some key gestural responses exhibited by cabinet officials Joseph Calton and former President Ronald Reagan, 1975

NPTEL

2010

So if you have a look at this nice picture you can see this with production of sentences, especially the punctuation this means the important syllables, the sentence stress and so on, this is highly synchronized in a temporal way with certain movements of my head, of course but also of my eyelids, eye closing and other aspects, in this case with the hands.

So I present this here with a posture because it is nice to also talk about not only self-coordination but coordination between people. So for example if people in a group refer to each other and are directed for to the same audience they also tend to synchronize in certain way.

This is not only for these kind of politicians in, depicted in this picture but also if you talk to each other on, on a table for example, you will observe, especially with the fast forward of your video that people tend to use the same positions, for example leaning on table, leaning forward, leaning backward.

This is not highly synchronized but it is a pattern that you can observe.

(Refer Slide Time: 11:32)

Multimodal Action

- Outline:**
- Introduction: Verbal and nonverbal information
 - Gestures
 - Posture
 - Gaze
 - Voice
 - Space
 - Turn taking
 - Emotion
 - Social relationship
 - Summary



Now I have to go back a little bit

(Refer Slide Time: 11:37)

Posture

- Coordination**
- Gestures, postures, facial expressions and
 - Self-synchrony
- Interaction synchrony (aligner)
- Interaction coordination
- Back-channeling
 - Turn-taking



to the last slide because I missed one point actually. So there is also interaction coordination but I will talk about this a little bit later this week when I talk about turn-taking and back-channeling.

Just to mention here, back-channeling for example are all these

(Refer Slide Time: 11:56)

Posture

Coordination

- Gestures, postures, facial expressions and
- Self-synchrony



FIGURE 7-14 | Photo of group conference exhibited by author Alfred Joseph Collins and former President Dwight D. Eisenhower, 1957. , 2010



kind of so-called recipient signals that we produce when we are actively listening to people, to a person talking.

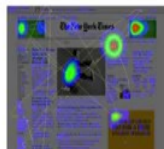
So for example, saying uh huh or nodding with the head, or smiling or something like this, all these signals which just show I am listening. I am still engaged in the conversation. You have my concentration. I am focused on you.

All these signals are of course, highly synchronized with the speech and the non-verbal information of the real talker who has the floor.

(Refer Slide Time: 12:27)

Gaze

- General: listeners look at speakers more often (60%) than the other way around (30%, Exline, 1974)
- Looking at the interlocutor is longer (e.g. 3 sec.) than eye contact (about 1 sec.)
- Gender: (US American) women look more at the conversational partner than men
- Likeability: people look more at likeable interlocutors than not likeable ones
- Status: prestigious / powerful people look not less when listening compared to speaking
- Cultural dependent (e.g. less gaze contact in U.S. than Europe)



So now coming to gaze, gaze you might have heard about this already is also important or well advanced in human computer interaction. Gaze signals interest and concentration. Which areas are now in the focus, the real focus of myself?

In this case, in the focus of my eyes. So here you have a nice photograph of so-called

(Refer Slide Time: 12:54)

Gaze

- General: listeners look at speakers more often (60%) than the other way around (30%, Exline, 1974)
 - Looking at the interlocutor is longer (e.g. 3 sec.) than eye contact (about 1 sec.)
- Gender: (US American) women look more at the conversational partner than men
- Likeability: people look more at likeable interlocutors than not likeable ones
- Status: prestigious / powerful people look *not* less when listening compared to speaking
- Cultural dependent (e.g. less gaze contact in U.S. than Europe)



heat map which just shows aggregated over time, the time of the focus of the eyes measured by eye tracking. We will talk about eye tracking in the second half of this lecture later on.

Gaze also signals not only interest but also other social aspects. For example, if I am talking, usually I gaze less to my dialog partner than when I am listening. Also social roles show differences in gazing. The gaze duration is typically a little bit increased for people with lower social roles.

And therefore they are for sometimes in some cultures, for example western cultures also gender differences that means that females on average tend to gaze a little bit more when they are listening towards the dialog partner who is speaking.

Also likeable people or attractive people are usually watched and gazed more. But be careful. Just looking in the general direction is not the same as actually focusing on a certain person in sense of making eye contact. Eye contact is really typically quite short because it can easily be considered as flirt.

Prestigious and highly self confident people tend to have a little bit different pattern. When they are talking, in contrast to the rest of the population they show even more increased gaze than when they are listening. So this difference between talking and listening is not as strong when highly popular or highly prestigious people are in focus.

All that been said, it is just an average. So you can observe this but typically there is a whole bunch of variance here. This means it is highly cultural dependent and situation dependent. And there are also a lot of individual differences here.

So you cannot easily come up or identify certain roles, status or interest just by looking at the gaze patterns of a person. But there are general average tendencies that you can observe. And of course it is interesting for human computer interaction to react on gazing.

For example consider a system that should support people in planning a travel. So if they is a couple or family coming to this device or a system in service and once you get information for nice flight or nice holidays, so it is of course important to, or could be beneficial to analyze the gaze patterns of these people in order to know when this system is targeted or addressed or when the people are engaged in discussing with each other the content of the system and the situation.

(Refer Slide Time: 16:03)

Voice

- **Social signals in voice** (Bolinger, 1964, Knapp & Hall, 2010)
 - Social dominance and self-confidence is often signaled with low or falling pitch, higher tempo and intensity; also frequent interruptions of other people
 - The opposite is observed for unoffending or needy people
 - Also cultural dependent
- Age, gender, attractiveness, emotional state, height, weight
- Personality (extraversion – introversion)
 - Extraversion: higher tempo, pitch variation, less pauses
- Education, intelligence
 - Socioclect, Regiolect etc.
- Pragmatic context (irony, question/statement)
- ➔ Although listener often rate consistently, only few ratings are correct
 - E.g. extraversion, age, gender, weight, context



Let us come to voice. As I said in the beginning of this week's lecture voice is used or expresses a lot of information revealing information of the person speaking, so for example age, gender, regional background, certain physical properties of weight for example and also attractiveness. But also emotions are exhibited in voice.

And even some aspects of personality. For example, highly extrovert people typically speak a little bit faster and have more pitch variation than others, than introvert people.

On the downside the problem is that we as human listeners tend to attribute a lot of aspects and characteristics towards people which we can see or people that, in this case we can hear although these might not be even be valid.

So other personality traits apart from extroversion and introversion might not be actually visible or perceivable in the acoustic information. Nevertheless, we tend to, humans tend to characterize people based on such information regardless their validity.

I also want to mention the pragmatic context. That means that in voice, aspects of irony or how to interpret the sentence, meaning is affected and influenced by the way we are speaking. So by the pitch pattern or by the way we are stressing the certain sentence or certain syllable in the sentence.

(Refer Slide Time: 17:58)

Multimodal Action

- Outline:**
- Introduction: Verbal and nonverbal information
 - Gestures
 - Posture
 - Gaze
 - Voice
 - Space
 - Turn taking
 - Emotion
 - Social relationship
 - Summary



So the last topic for this video is space.

(Refer Slide Time: 18:05)

Space

Social signals in groups (Kendon, 2009)

- People organize themselves spatially
- Signaled by orientation of head (and upper body, sometime whole body)
- Shared transactional space (F-formation)
 - inner space
 - participant's space
 - immediate environment



Of course with the emergence of social robots or other autonomous systems it is highly important to know how people behave in space and how they orient each other and use space and manage to coordinate themselves in a given area.

Here you see the typical so-called F-formation. I do not know actually what F-Formation means but I am, I would say that it might mean the free field formation, or the free formation, that means how people orient themselves in a given free space when you do not have, like chairs or other obstacles that restrict people how they place themselves in the space. If you are interested, try to find it out, do some research what F-Formation is based on.

What you can see here are actually 3 parts of the space. There is always, if you have a group, a small group or even a couple of people means two people engaged in a conversation or acting as a group, they do not have to talk, during this time they just have to know or feel that they are behaving like a group, you have the so-called O space in the center.

This is the interaction space, the space which is not occupied and we need this to see what the other people are doing, to recognize the face expressions, to recognize the posture, to see the gesturing and also to really good, really well hear what the people are talking. So this is the interaction space which is not occupied for transmitting social and verbal signals.

Then we have P, the private space. This is where the people are actually arranged. The private space is quite small and it covers not only the persons acting, but also the personal belongings

like may be the jacket or the bag. Then we have the third kind of space, the outer space which is not part of this kind of interaction.

(Refer Slide Time: 20:22)

Space

Social signals in groups (Kendon, 2009)

- You have to be invited into p
- 1. Display interest
- 2. Group acknowledgement
- 3. Invitation by offering space



Forgan, 1999



So if there is a group talking to each other in this so-called F-formation, engaging or coming into this group is usually done in a so-called, in a certain manner. That means you cannot just interrupt there, go there into the private space and start interacting.

You have to be invited. This means you usually try to coming from the immediate environment to catch some attention and then being invited to join the private space. People will look at you; make space in this private space area for you. Here we see the three phases of that.

(Refer Slide Time: 21:09)

Space

- In dyadic interaction, there are two main formations:
- Vis-a-vis: greeting, relationship is topic (right)
- L: talking about a topic (left)
- Often shift from left to right
- Alternative: side-by-side
Topic: immediate environment



But of course there are other formations, not only the, Vis-a-vis position where people are actually in front of each other but they can also have different angles. It really depends on the certain situation what they are talking about and what is the surrounding environment, how they place each other.

For example, in a given conversation these formations may change a little bit. So while we are focused on each other because we are just greeting each other or saying good bye then we may have a lot of attention towards each other, that means face-to-face vis-a-vis and then after a while it will be little bit too intensive so we may just have certain angle to each other like 90 degrees or something like this.

And even people who know each other very well or watching something else might, be like, like an old couple, they might stand side by side and watching the immediate environment while talking to each other.

It would be awkward if this system which is moving would be too close or entering the private space of people. But if you want to have an interactive system it would be also awkward to be too far away.

So knowing these formations and how to actually interact into engaging in a conversation or an interaction is important for such mobile systems.

And again we have highly cultural dependencies. That means for certain cultures this O space or P space might be different than for others.

(Refer Slide Time: 23:50)

Space

Example space (e.g. environment affects interaction)
• Bar

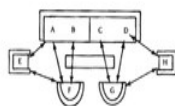


FIGURE 4-6
Concentration flow and furniture arrangement.
Source: From Public Places and Private Spaces, by Albert Middelton. Copyright © 1978 by Van Nostrand Reinhold, Inc. Reprinted by permission of Van Nostrand Reinhold, a member of Pearson Education, L.L.C.

Taken from Knapp & Hall, 2010



I now want to close the space aspect with two examples for more straight formations, not the F-formations nor the free formation but where the environment actually shapes the kind of conversation or interaction that is going on.

And a bar for example, think about this scheme or this figure as bigger bar or cocktail bar with chairs where we really sink in. You would not just stand up and move the chairs around. They are really heavy and fixed. But this formation, this chair layout actually influences who is talking to whom.

Usually

(Refer Slide Time: 24:32)

Space

Example space (e.g. environment affects interaction)

- Bar

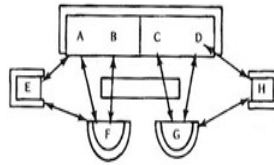


FIGURE 4-6

Conversation flow and furniture arrangement.

Source: From *Public Places and Private Spaces*, by Albert Mehrabian. Copyright © 1976 by Basic Books, Inc. Reprinted by permission of Basic Books, a member of Perseus Books, L.L.C.



Taken from Knapp & Hall, 2010

you really need some kind of easy eye contact. That means you are rather not talked to the person sitting right next to you but the person you really do not have to orient to without any difficulties. So the arrows here are indicating the most likely conversations that are going to happen when people talk to each other.

(Refer Slide Time: 24:59)

Space

Example space (e.g. environment affects interaction)

- Bar

- Class-room

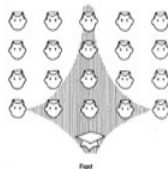


FIGURE 4-7

The size of the participant.

Taken from Knapp & Hall, 2010



The second example is the classroom and this is just indicating the kind of focus I would have as a teacher or lecturer while people are in a classroom or students in the audience hall. This is just the focus of the concentration and attention that a person speaking to bunch of people in chairs would have. So again the environment also shapes the way people interact with each other.

(Refer Slide Time: 25:31)

Space

Example space (e.g. posture)

- Exact dimensions of personal space depend on
- Culture (typical private and inner space)
- Context (formal vs. private)
- Status (powerful persons get more space, but might initiate smaller distance)



As a summary space is highly cultural dependent and context dependent. There is a real difference between private and more public situations and relationships.

Again an old couple standing side by side talking to each other would be perceived most likely as highly familiar and intimate where as people who do not know each other that well, like business partners. If they would do that it would indicate no familiarity or intimacy but rather may be having a problem with each other.

Again status and social role also affects the space difference. So not only the cultural differences but also the power, social power affects how close we get to each other.