

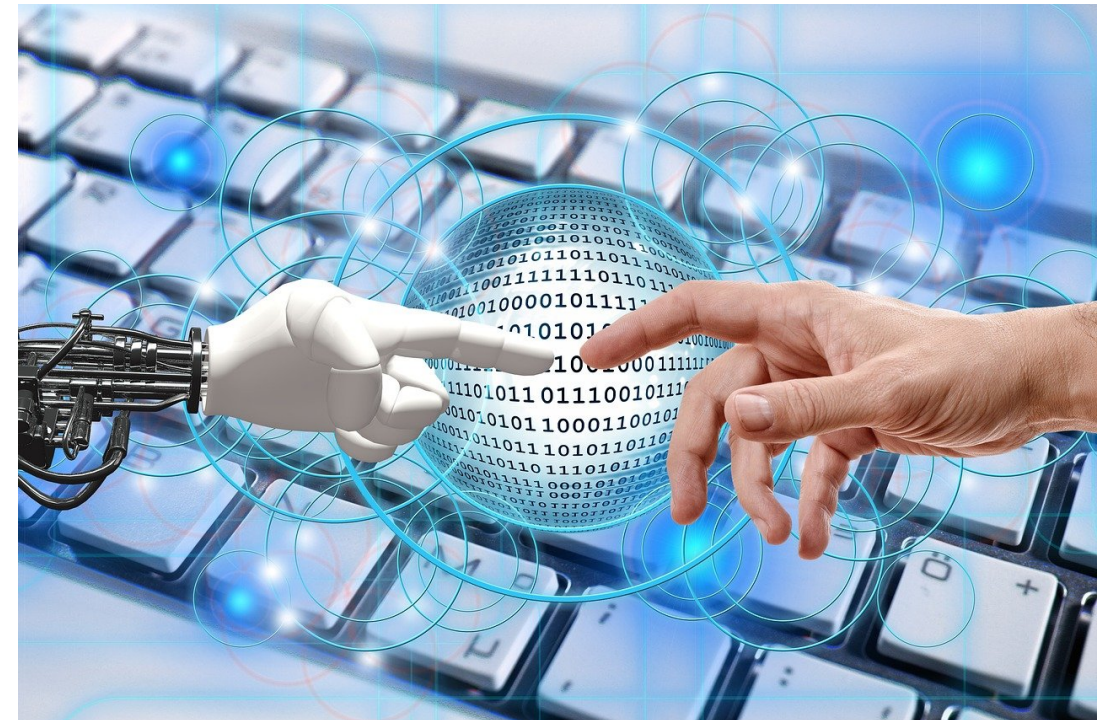
Lecture 04 - Verbal Communication in Humans and Robots

Prof. Dr. Teena Hassan
teena.hassan@h-brs.de

Department of Computer Science

Hochschule Bonn-Rhein-Sieg
Sankt Augustin

2nd May 2024



- What did we learn in the previous lecture about nonverbal communication?
- 2 minutes to note down your points
- DO NOT look at the slides.

Participants can vote at [Slido.com](#) with [#1316882](#)
and the passcode: on93mx (26 Oct) or anytime at [this link](#)

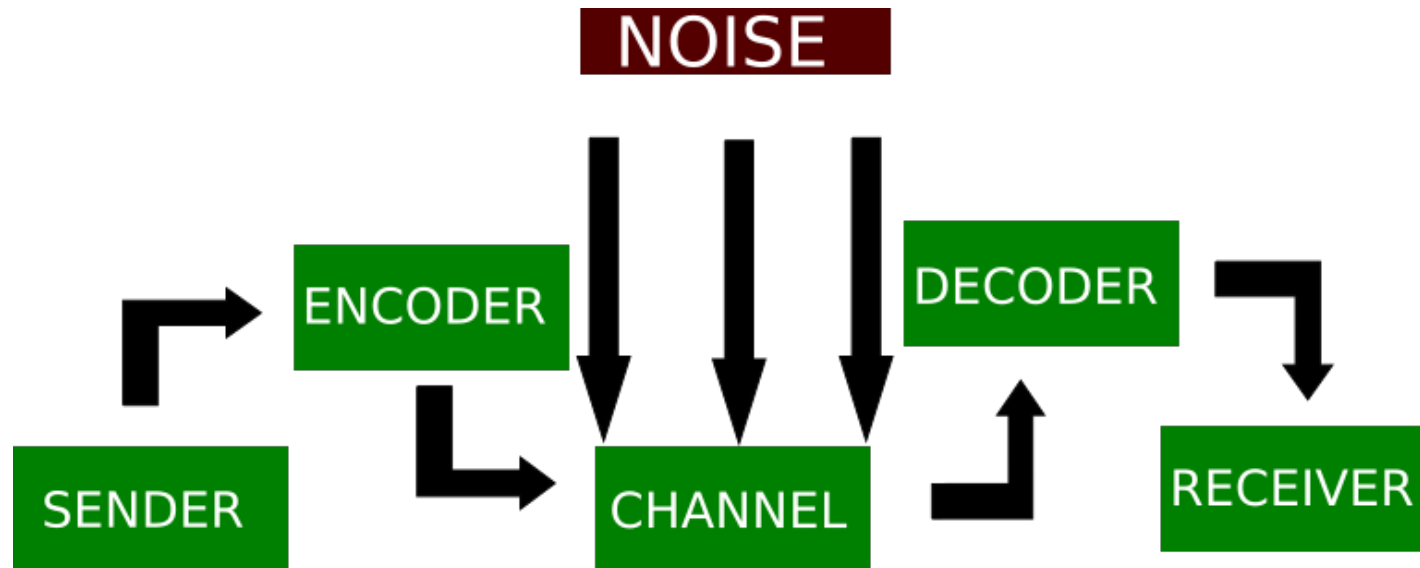
- At the end of today's lecture, you will be able to:
 1. Describe models for communication between humans.
 2. Identify the key elements and concepts of verbal communication.
 3. Elucidate the technical aspects of realizing verbal communication between humans and robots.



<https://youtu.be/dBT6u0FyKnc>

- Obviously not!
- Researchers have looked at how humans communicate, in order to identify:
 - The **elements** of communication
 - The **modes** of communication
 - The **processes** involved in communication
 - The **factors** that influence communication
- .. and to create **a model for interhuman communication and comprehension...**
- ... which could be used to study communication problems as well as to design artificial communicative agents.

- Transport the message from source to destination
- Information **sender** – source that produces the message.
- Information **receiver** – destination for whom/which the message is intended.
- **Channel** – medium that carries the message.
- **Encoder** – converts the message into a signal that can be transmitted through the channel.
- Channel introduces **noise** in the transmitted signal.
- **Decoder** – converts the received noisy signal back into the message.



Shannon, C.E., Weaver, W.: A Mathematical Model of Communication. University of Illinois Press, Urbana, IL (1949)

- The linear model is not enough.
- **Context** of communication is important.
- Communication is **multimodal**.
 - Not only: What has been said?
 - But also: How was it said?
- **Feedback loops** between senders and receivers:
 - Communication is a cyclic process.
 - Reciprocity is an important part of social norms and polite behavior.
 - ▶ A greeting is expected to be returned; a question is expected to be answered, etc.
- Sender and receiver **co-construct** the meaning of communication.
 - Coding and decoding are dependent on each other.

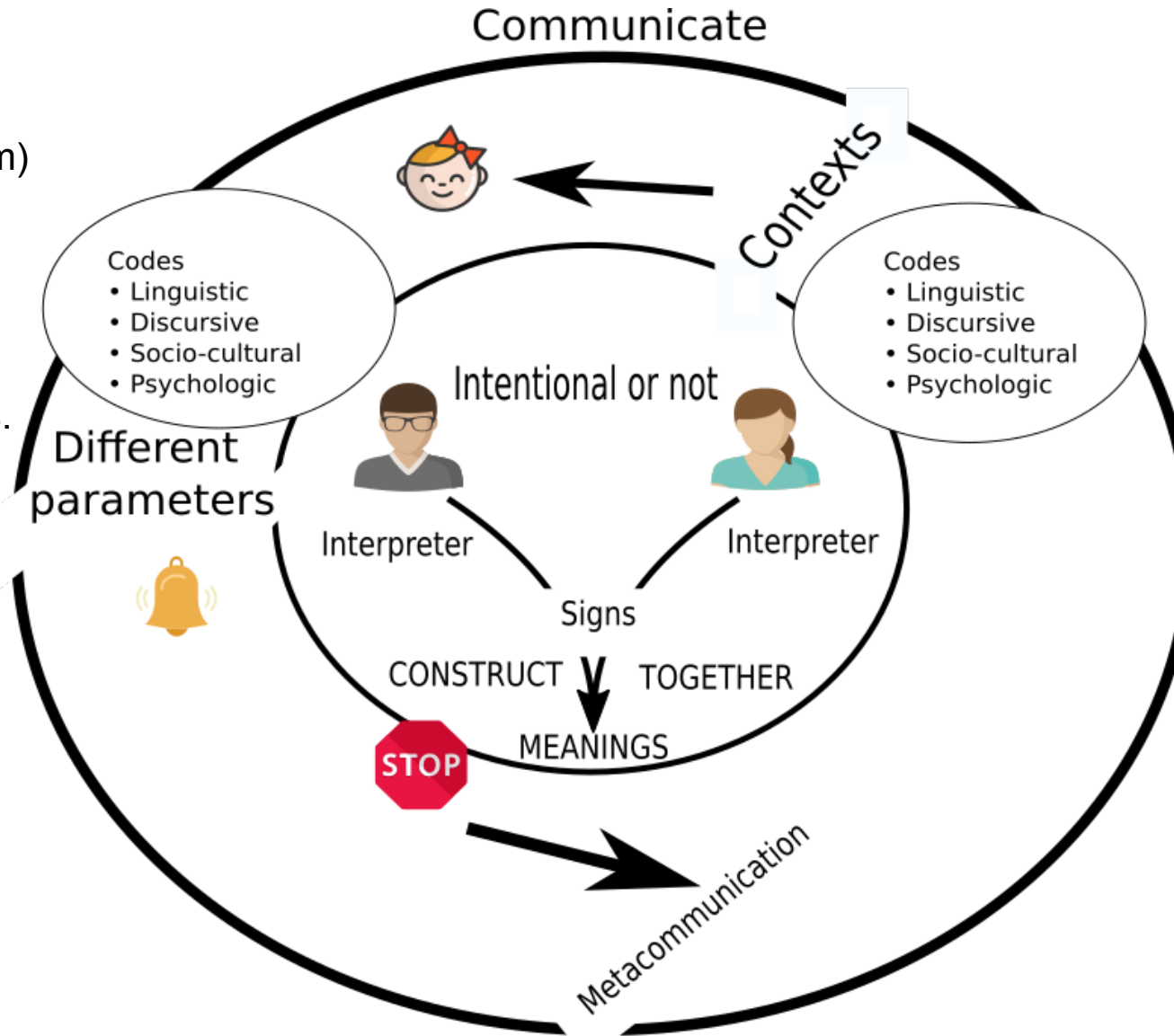
We "cannot not communicate."
-- (Watzlawick et al., 1967)

- **Communication context**
 - Temporal (when)
 - Spatial (where)
 - Socio-cultural (with whom)
 - Others
 - Objects
 - Audience

- Same message has different meanings in different contexts.

Intentional or not, we are **continuously communicating**.

- Others are perceiving and interpreting us, even if we do not intentionally express signs.



Each individual uses **codes** to express as well as interpret communicative signs.

- Our internal representation (model) of the world is key here.

Metacommunication: Communication about communication.

- e.g. clarification requests

- **Can be applied to human-robot communication.**

Robots should be aware that the human is constantly interpreting the signs sent or not sent by the robot and this could influence the subsequent interaction with the robot.

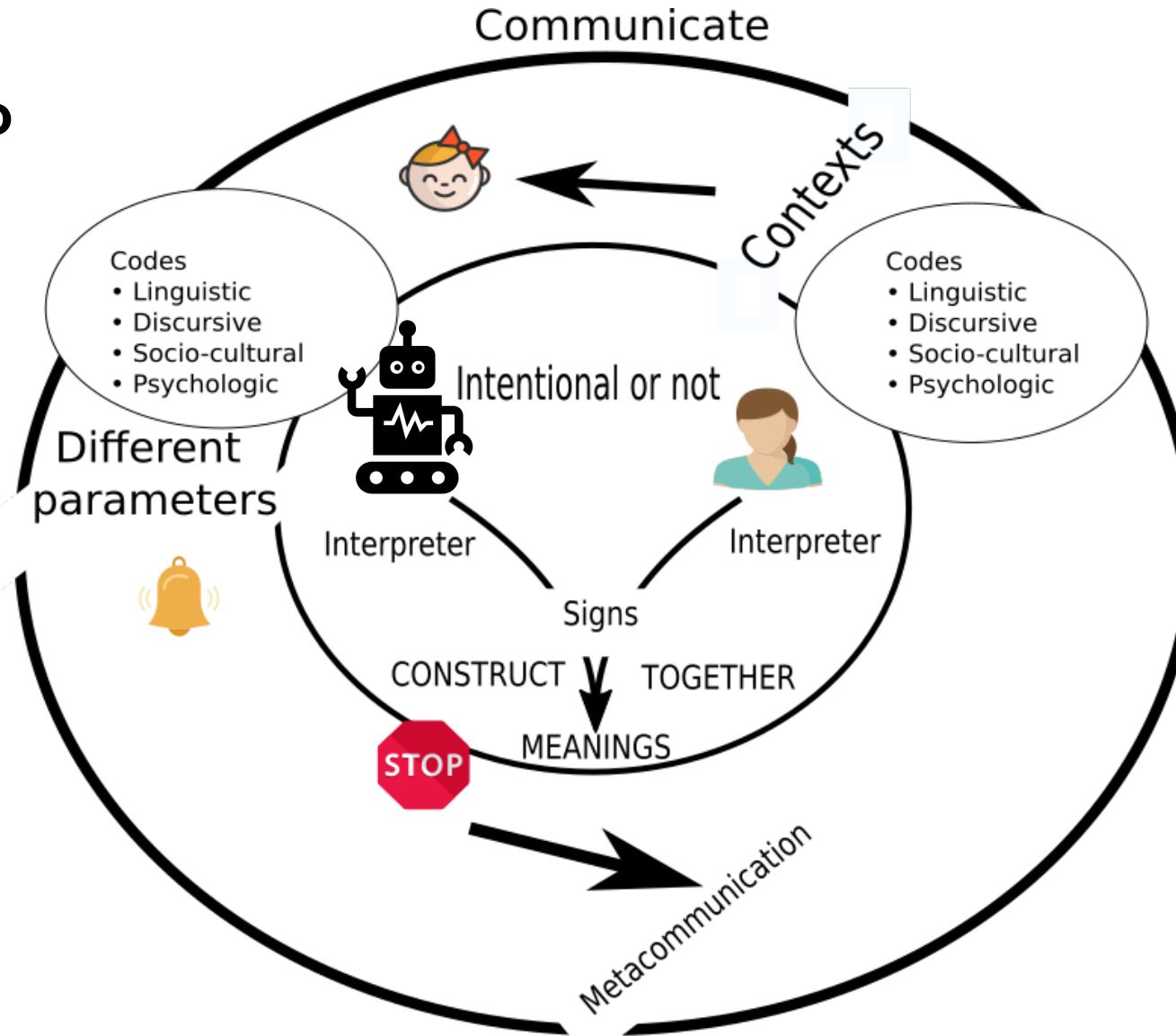
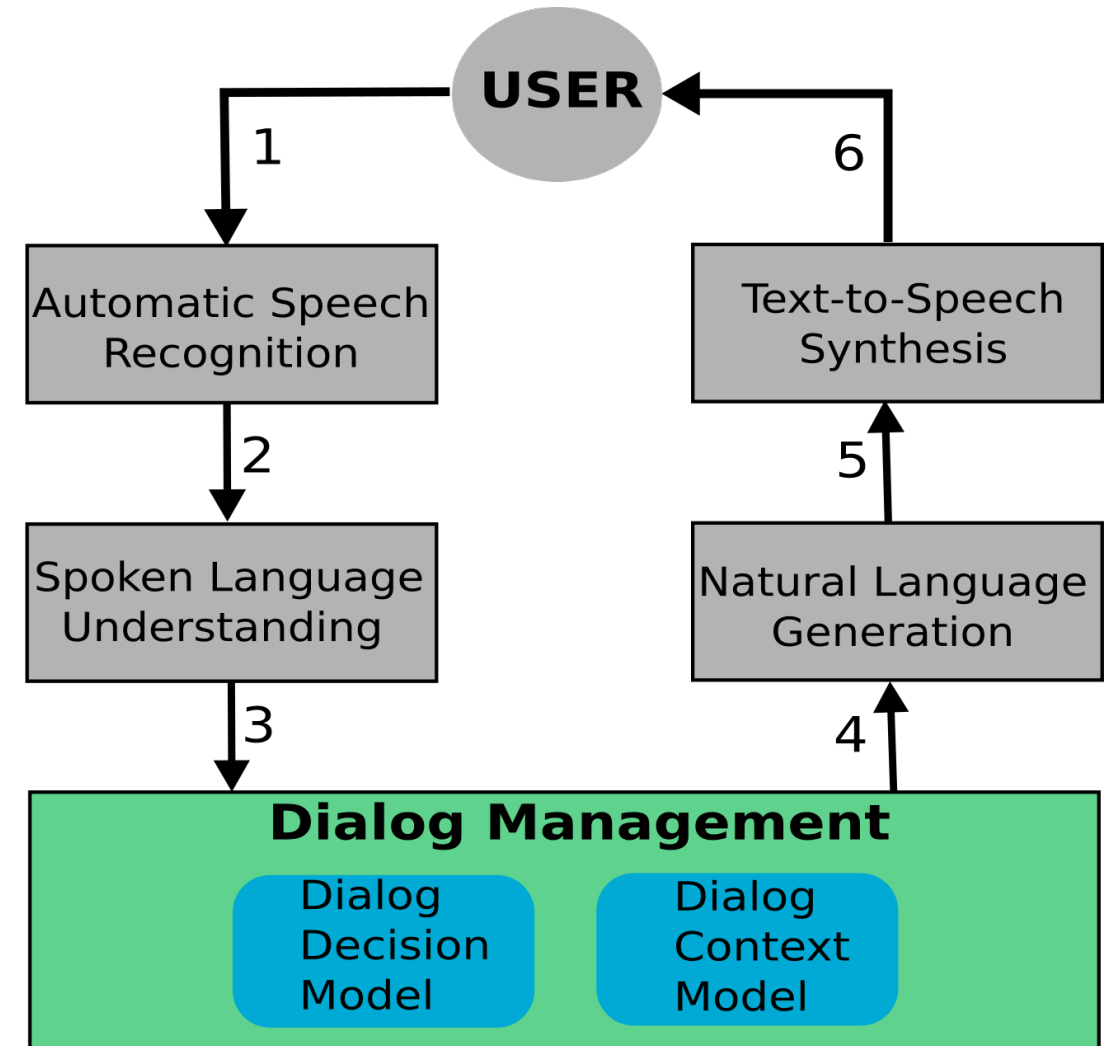


Image recreated from (Grandgeorge, 2020) and (Blanchet, 2000)

- **Verbal communication** *"is the ability of humans to express their thoughts and to communicate with each other through a system of vocal and/or graphic signs constituting a language."* -- CNRS, (Grandgeorge, 2020)
- **Language** is *"a system of communication by speaking, writing, or making signs in a way that can be understood, that allows the mention of (not only) present but also past and future situations."* -- (Grandgeorge, 2020)
 - A system which consists of a set of symbols (sentences) — realised phonetically by sounds — which are used in a regular order to convey a certain meaning. <https://www.uni-due.de/ELE/LinguisticGlossary.html>
- Spoken language:
 - Input: Auditory channel
 - Output: Vocal channel
 - Basic unit: Phonemes
- Sign language:
 - Input-Output: Visual-bodily channels
 - Formational units:
 - ▶ hand shapes
 - ▶ locations
 - ▶ movements
 - ▶ orientations
 - ▶ non-manual

- The processing pipeline and components needed for realizing spoken dialog systems.
 - Here, we are looking only at the verbal part of communication.
 - If the dialog system is used together with an embodiment (virtual or physical), then nonverbal communication is inevitable.
 - ▶ A full-fledged multimodal interaction architecture would be needed.



Recreation of Fig. 1 in (Cañas et al., 2021)

- Form 5 groups
- Assign one component to each group
 - ASR / NLU / Dialog management / NLG / TTS
- Search for papers, blog articles, source code repositories and videos describing different approaches and methods for these components
- 15 minutes to search
- 5 minutes to document
- 10 minutes to present
- <https://app.conceptboard.com/board/0idn-ud98-peht-0ofx-u76k>

1. User makes a query, a command, etc. in the form of an **acoustic signal**.
2. **Automatic Speech Recognition (ASR)** analyses the acoustic signal to recognise the spoken words (along with an estimate of the agent's confidence in the recognised words).
3. **Spoken (Natural) Language Understanding (NLU / SLU)** analyses the recognized words and extracts the intents (instruct, query, reply, greet, etc.) and entities (name, place, time, thing, etc.) relevant for the application domain.
 - "Hello Robbie!" --> "Hello __name__" --> The user is greeting and Robbie is a name.
4. **Dialog Management (DM)** consists of two models:
 1. **Dialog Context Model (DCM)**
 - ▶ Tracks the dialog state and is updated by DM with information provided by NLU.
 2. **Dialog Policy Model (DPM)**
 - ▶ Decides the agent's next dialog act (intent, entities) given the information in DCM.
5. **Natural Language Generation (NLG)** converts the agent's dialog act into a sequence of words.
6. **Text to Speech (TTS)** converts the words into acoustic signals which are then heard by the user, and which could trigger another round of exchange.

- Who controls the conversational floor?
 - That is, who would speak at a specific point in time. Consequently, others listen.

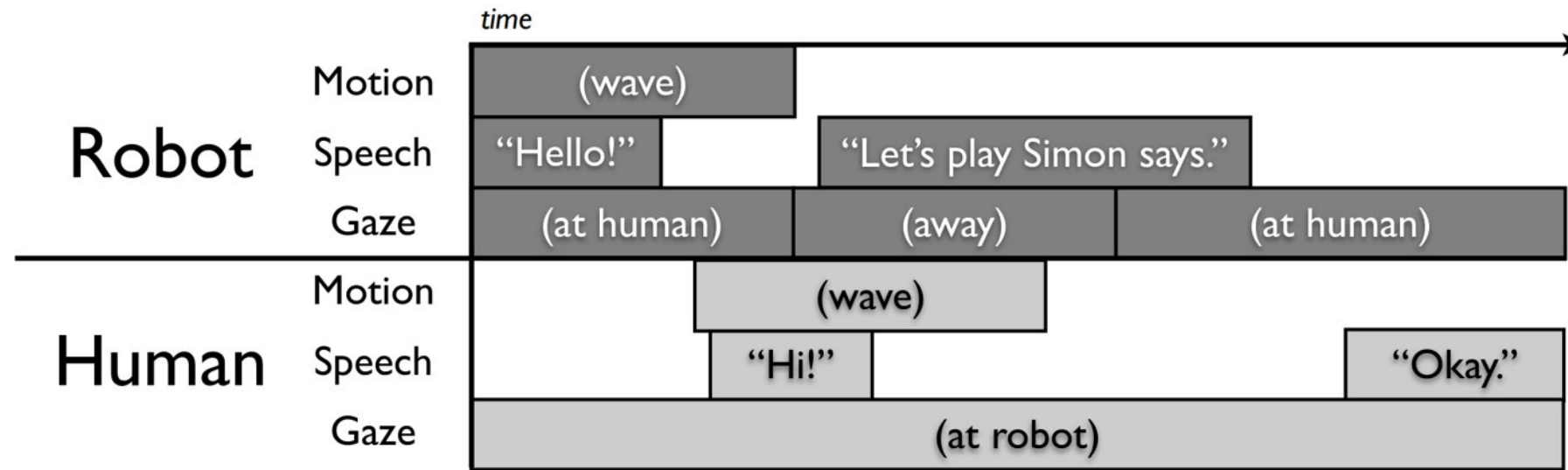


Figure 1. An example of event alignment for a multimodal interaction.

Source: Figure 1 from (Chao and Thomaz, 2012)

- (Chao and Thomaz 2013) proposed Timed Petri Nets (TPN) for dynamic turn-taking during one-on-one interactions between a human and a robot.
 - CADENCE Architecture
 - Four parts of floor regulation during conversations:
 - ▶ seize the floor
 - ▶ yield the floor
 - ▶ hold the floor
 - ▶ audit the owner
 - **Backchanneling** from listener to speaker to confirm that the listener is paying attention to the speaker, is in agreement with the speaker, etc.
 - ▶ Through verbal (yeah, mhm, uh-huh, etc.) or nonverbal cues (head nods, facial expressions, etc.).
 - ▶ Speaker holds the floor.
 - Backchanneling could also be used by listener to (eventually) seize the floor from the speaker.
 - ▶ E.g. "Yeah, but"
- Part of Dialog Management component of spoken dialog systems.

- At a dining table, PersonA says to PersonB: "*Could you please pass me **the salt?***"
 - PersonB would know that PersonA is referring to **the salt on the table** next to PersonB.



- In **situated interaction**, an agent should have awareness about **spatial and temporal relations** from own perspective and other's perspective, and...
- ... should be able to **connect symbols with their own sensorimotor flow** (i.e. with what they are currently sensing and doing).

- What symbols should the robot know and what do they mean?
- How can it learn new symbols and their meanings? (*symbol emergence*)
- Human-centered aspect: How to adapt the vocabulary of the robot to the human interaction partner?

- Some of the symbols (words) in the vocabulary will be linked to the physical environment.
- The robot should be able to distinguish whether the symbols referred to by the human (or the other agent) are abstract concepts or actually things in its current environment.
- It should then identify that thing in its environment and link the symbol to it (*symbol grounding*).
 - Internal representation of the physical world.
 - ▶ Spatial and temporal relations.
 - ▶ Updated based on sensorimotor flow.
 - Perspective-taking

Group Activity — 20 + 10 Mins



Write a script for a conversation between a human and Pepper robot in a museum setting. The robot is supposed to guide groups of visitors through the AI exhibition hall and give information about the AI exhibits in the room. During the design of the script, include any THREE of the following concepts associated with verbal communication and highlight how you included it in the script:

Turn-taking

Backchanneling

Metacommunication

Symbol grounding

Such types of applied questions could be asked in the exam.

- In today's lecture, you learnt to:
 1. Describe models for communication between humans.
 - ▶ Blanchet's ethno-sociolinguistic model for human communication
 2. Identify the key elements and concepts of verbal communication.
 - ▶ Verbal: Spoken language, sign language, turn taking, symbol grounding
 3. Elucidate the technical aspects of realizing verbal communication between humans and robots.
 - ▶ Architecture of spoken dialog systems

- Grandgeorge, M. (2020). **Communication Between Humans: Towards an Interdisciplinary Model of Intercomprehension**. In: , *et al.* Human-Robot Interaction. Springer Series on Bio- and Neurosystems, vol 12. Springer, Cham. https://doi.org/10.1007/978-3-030-42307-0_1
- Doncieux, S., Chatila, R., Straube, S. *et al.* Human-centered AI and robotics. *AI Perspect* **4**, 1 (2022). <https://doi.org/10.1186/s42467-021-00014-x>

[References 1 would be important for the exam.]

- **Blanchet**, P.: Linguistique de terrain, méthode et théorie (une approche ethnosociolinguistique), 150 p. Presses Universitaires de Rennes, Rennes (2000).
- Pablo **Cañas**, David Griol, Zoraida Callejas, "Towards versatile conversations with data-driven dialog management and its integration in commercial platforms", Journal of Computational Science, Volume 55, 2021.
- Crystal **Chao** and Andrea L. Thomaz. 2013. Controlling social dynamics with a parametrized model of floor regulation. J. Hum.-Robot Interact. 2, 1 (February 2013), 4–29. <https://doi.org/10.5898/JHRI.2.1.Chao>
- C. **Chao** and A. L. Thomaz, "Timing in multimodal reciprocal interactions: control and analysis using timed Petri nets," Journal of Human-Robot Interaction, Vol. 1 No. 1, 2012, 78-95.
- **Doncieux**, S., Chatila, R., Straube, S. et al. Human-centered AI and robotics. AI Perspect 4, 1 (2022). <https://doi.org/10.1186/s42467-021-00014-x>
- **Grandgeorge**, M. (2020). Communication Between Humans: Towards an Interdisciplinary Model of Intercomprehension. In: Jost, et al. (Eds.) Human-Robot Interaction. Springer Series on Bio- and Neurosystems, vol 12. Springer, Cham. https://doi.org/10.1007/978-3-030-42307-0_1
- **Shannon**, C.E., Weaver, W.: A Mathematical Model of Communication. University of Illinois Press, Urbana, IL (1949).
- **Watzlawick**, P., Beavin-Bavelas, J., Jackson, D.: Some tentative axioms of communication. In: Pragmatics of Human Communication: A Study of Interactional Patterns, Pathologies and Paradoxes. W. W. Norton, New York (1967)

- 09.05.2024: Public holiday, no lecture and tutorial.
- Lecture on "Building socially interactive agents — components and examples"
 - Thursday, 16.05.2024 at 9 am
 - In: B060, Grantham Allee 20, 53757 Sankt Augustin

ADDITIONAL SLIDES



- Each interpreter interprets the signs based on their own codes.
- When the received responses diverge from the intended outcome, then it indicates a discrepancy in the ongoing communication.
- The communication could be interrupted by at least one of the interpreters and they could now discuss about the communication itself.
- This is metacommunication.
- After the resolution of the misunderstandings, the original exchange can be resumed.
- Metacommunication involves high-level cooperation between the agents, where the interpreters are able to analyze and discuss about their own communication strategies.

- "Meanings 'given off' (Goffman) by an individual's body language through nonverbal leakage, or unconsciously signified by their appearance, dress, or behaviour, including whatever may be noticeable by its absence in a particular context (see *also* analogic communication). As Watzlawick observed, 'one cannot *not* communicate', regardless of whether an observer's inferences are warranted."
 - Oxford Reference (<https://www.oxfordreference.com/view/10.1093/oi/authority.20110803110707314>)

- **ethnography of communication** “The study of cultural differences in acts of communication. This is a comprehensive term which goes beyond simple differences in language to cover additional aspects such as formulaic use of language (e.g. in greeting or parting rituals), proxemics (the use of distance between partners in a conversation) and kinesics (the study of body movements used in communication).”
- **sociolinguistics** “The study of the use of language in society. Although some writers on language had recognised the importance of social factors in linguistic behaviour it was not until the 1960's with the seminal work of Labov that the attention of large numbers of linguists was focussed on language use in a social context. In particular the successful explanation of many instances of language change helped to establish sociolinguistics as an independent sub-discipline in linguistics and led to a great impetus for research in this area.”
- Source: <https://www.uni-due.de/ELE/LinguisticGlossary.html#GlossF>

- **Phoneme:**

- *“In traditional phonology the smallest unit in language which distinguishes meaning, e.g /k/ and /g/ as seen in coat and goat. Each phoneme has one or more realisations, called allophones.”*

- **Phonology:**

- *“The study of the sound system of one or more languages. Phonology involves the classification of sounds and a description of the interrelationship of the elements on a systematic level.”*

Source: <https://www.uni-due.de/ELE/LinguisticGlossary.html#GlossF>

- **Speech**

- *“The production of sounds using the organs of speech; contrasts directly with writing which is a secondary medium for communication via language.”*

- **Prosody**

- *“A term which refers to all the suprasegmental properties of language such as pitch, loudness, tempo and rhythm.”*
 - ▶ The nonverbal part of speech.

- **Suprasegmental**

- *“A reference to phenomena which do not belong to the sound segments of language but which typically are spread over several segments, e.g. intonation, stress, tempo, etc.”*

| | | |
|---------|--|--|
| request | instruct query propose check | task, cancel, backtrack open, select, yn suggest, offer align, confirmation, repeat |
| assert | provide | elaboration, statement, opinion |
| respond | yes no reply notify | agree, accept disagree, reject open, select acknowledge, buy_time, success, failure |
| social | greetings politeness interpersonal | opening, closing apology, thanks, acknowledge_thanks feedback |
| other | other | other_intent, no_intent |
| add_on | add_on | goes_with, correct |

Table 1: Dialog acts tags inventory.

Source: Table 1 in Silvia Pareti and Tatiana Lando. 2018. [Dialog Intent Structure: A Hierarchical Schema of Linked Dialog Acts](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).