



MAJOR TRENDS IN CONVERSATIONAL APPLICATIONS

MULTIMODAL CONVERSATIONAL AGENTS

DR. NAVA SHAKED,
CEO, BBT & HOLON INSTITUTE OF TECHNOLOGY AFFILIATION

Moderator: Moderator: Lorin Wilde, Wilder Communications

Nava Shaked - Bio

- Ph.D. in Computational Linguistics from CUNY GC
- Faculty member in Engineering Faculty in HIT, Israel
- R&D in speech processing in Bell Labs, IBM & more
- Vast Industry experience: Speech platforms, VUI Dialog design, Call Center technologies and Multimodality
- BBT Ltd.- NLP consulting for startups and enterprises
- AVIOS board and Israel's chapter chairman

DE GRUYTER

Nava Shaked, Ute Winter (Eds.)

DESIGN OF MULTIMODAL MOBILE INTERFACES

Nava Shaked, Ute Winter (Eds.)
DESIGN OF MULTIMODAL MOBILE INTERFACES

Design of Multimodal Mobile Interfaces discusses the fast developing new world of mobile multimodality, focusing on innovative technologies and design to create state-of-the-art user interfaces. It examines practical challenges entailed in meeting commercial deployment goals and offers new approaches to interface design. The “smart mobile” has become a powerful tool enabling us to perform multiple tasks. The user plays an important role in the mode of operation, so multimodal interaction provides the user with new complex multiple modalities of interfacing such as: speech, touch, type, gesture and more.

This book addresses the challenge of designing multimodal interfaces for mobile devices in a synergetic fashion that will not burden the user or create a technological overload. Each chapter addresses a different area of application: automotive, wearables, avatars & virtual agents. In addition it addresses expert modules such as NLP, Omni-channel, information classification & organization, and personalization modules.

- ▶ Discusses central questions concerning usability, concepts and complexity of mobile multimodality.
- ▶ Highlights current challenges and developments in multimodal interface technologies.
- ▶ Integrates various perspectives on technology, including future user and designer concerns and requirements.



Nava Shaked
Is a Professor at HIT, Holon Institute of Technology and head of the Multimedia, Multimodality Human Machine Interaction Center in Israel. She has 20 years of industry experience in the area of NLP & User interaction technologies.



Ute Winter
Is a Staff Researcher in the User Experience Technologies Group at General Motors Advanced Technical Center in Israel. She has more than 15 years of experience in Natural Language Processing and Human Machine Interface Technologies.



www.degruyter.com
ISBN 978-1-5015-1084-7



Virtual Agents & Avatars

AV&V are an AI based technology, enabling human machine communication. The “intelligence” is based on a wide range of knowledge and the ability to conduct Q&A interaction yielding a human like natural conversation.

Virtual Agent – as a customer service rep

A personal assistance for everyday tasks (Siri)

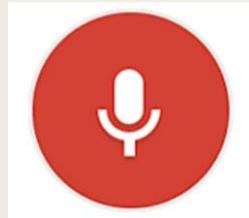
Siri



Microsoft
Cortana



Google
Voice Search



Amazon
Echo



Nina
Customer service



AVA
Airport Virtual Assistant



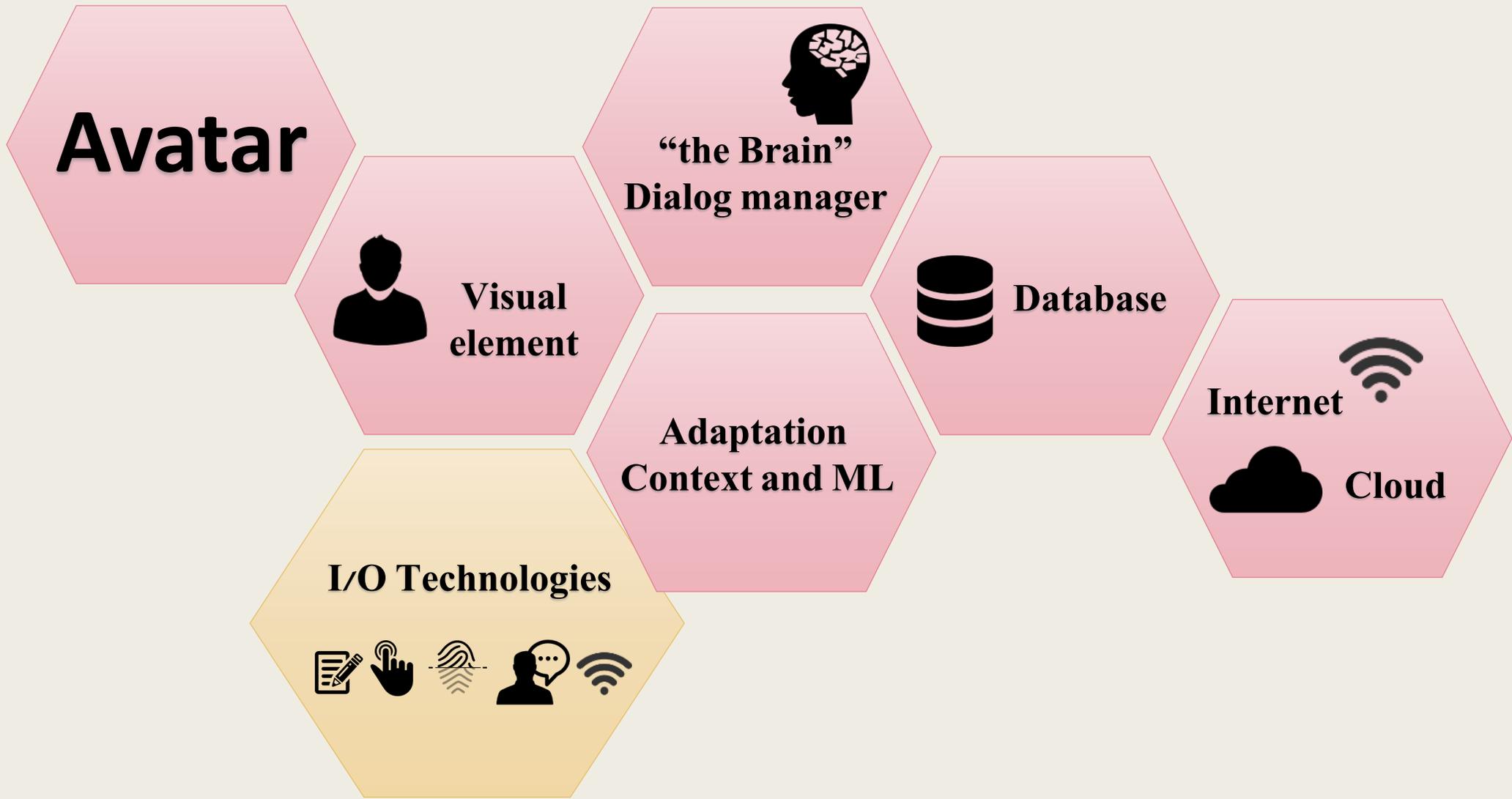
Avatars

Avatar is a virtualized personal implemented on different platforms. It can communicate in an intelligent manner and is usually an extension of the user self used to perform his tasks in games, Augmented reality, health applications etc.

In order to create A&VA we need to:

- Animation, graphics and 3D
- Investigate areas such as cognitive psychology, communication and human emotions and body movements for interaction





The A&VA Basic Modules



So what is so special about Avatars & Virtual Agents (A&VAs)?

- Casswell (2001) - Embodied Conversational Agents (ECA)

Translated the rules of engagement into a table of conversational functions and their behavior realization

This paradigm could be used for ECA-based systems so that the user interface will be as similar as possible to human-to-human interaction.



A&VA - Development stages:

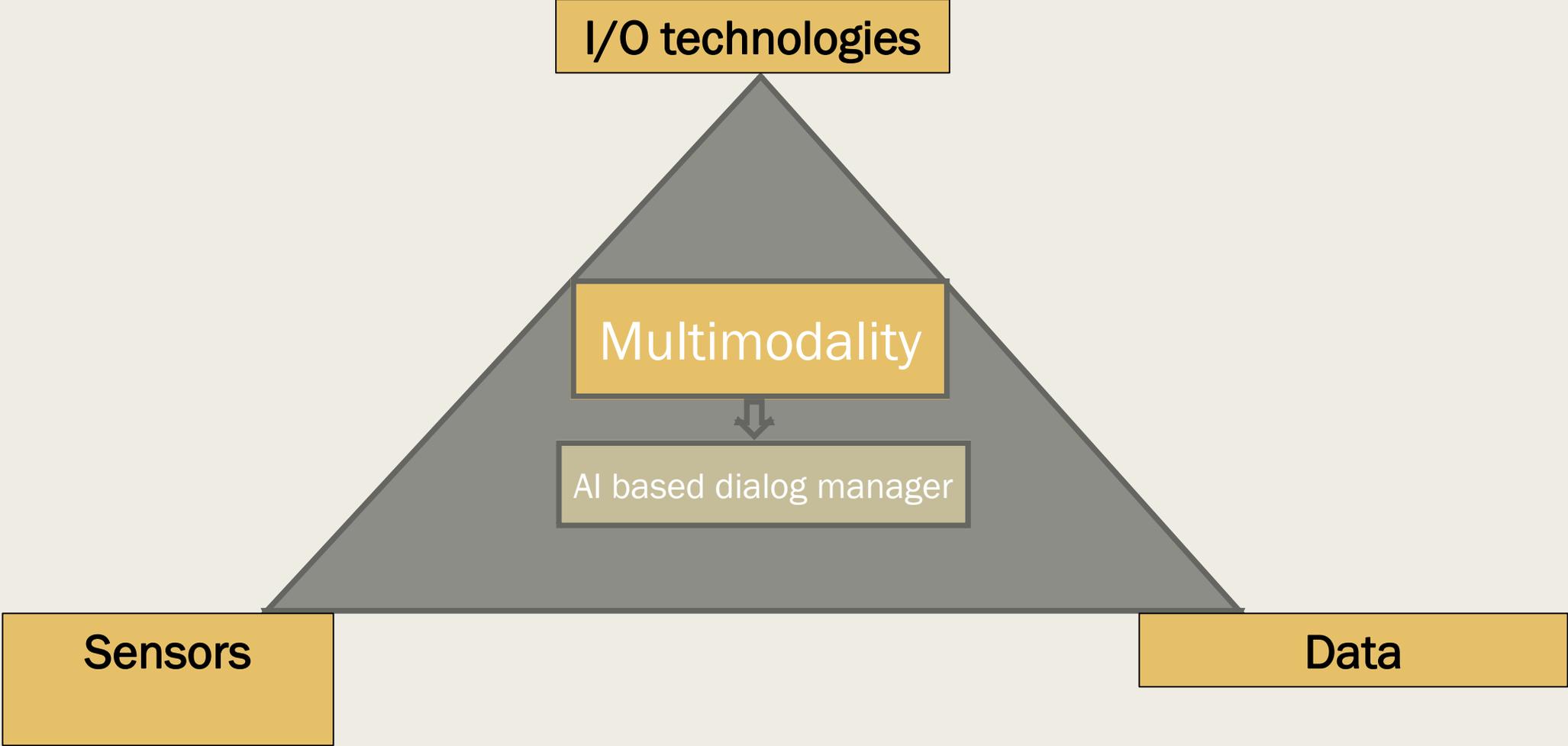


- 50's and '60's – Turing test - ELIZA, created by Stuart Weizenbaum
- 80's and '90's - the voice-based Virtual Agents in call center IVRs for self-service
- 90's as the internet was released for public use - Chat-boxes/chat bots used for customer service
- 2000's - Design/technology advancements - Digital avatars as virtual representations of real humans are also used in virtual worlds
- MMORPG (Massive Multiplayer Online Role Playing Game), the gamers choose a clearly combat focused Avatar
- 21th century – the clever species Medical Avatars, personal Assistance and service & sales agents.

The Change: leading to Multimodality

- The ability of an Avatar to generate face-to-face communication between real and virtual persons allows a much richer communication channel.
- It enables multimodal communication through both verbal and nonverbal channels such as gaze, gesture, spoken intonation and body posture coupled with sensors and real time data access and processing
- People create their own “virtual me” and “feed” it with multimodal data from sensors like fitness trackers or apps, so that they can measure weight loss, health or even sleep phases.

The interaction between humans and machines has reached a new level



A relationship framework for Avatars & Virtual Agents



Visual



Hearing



Voice



Touch



Text



Biometric



Gesture



WIFI



Speech



Camera



Touch



Text



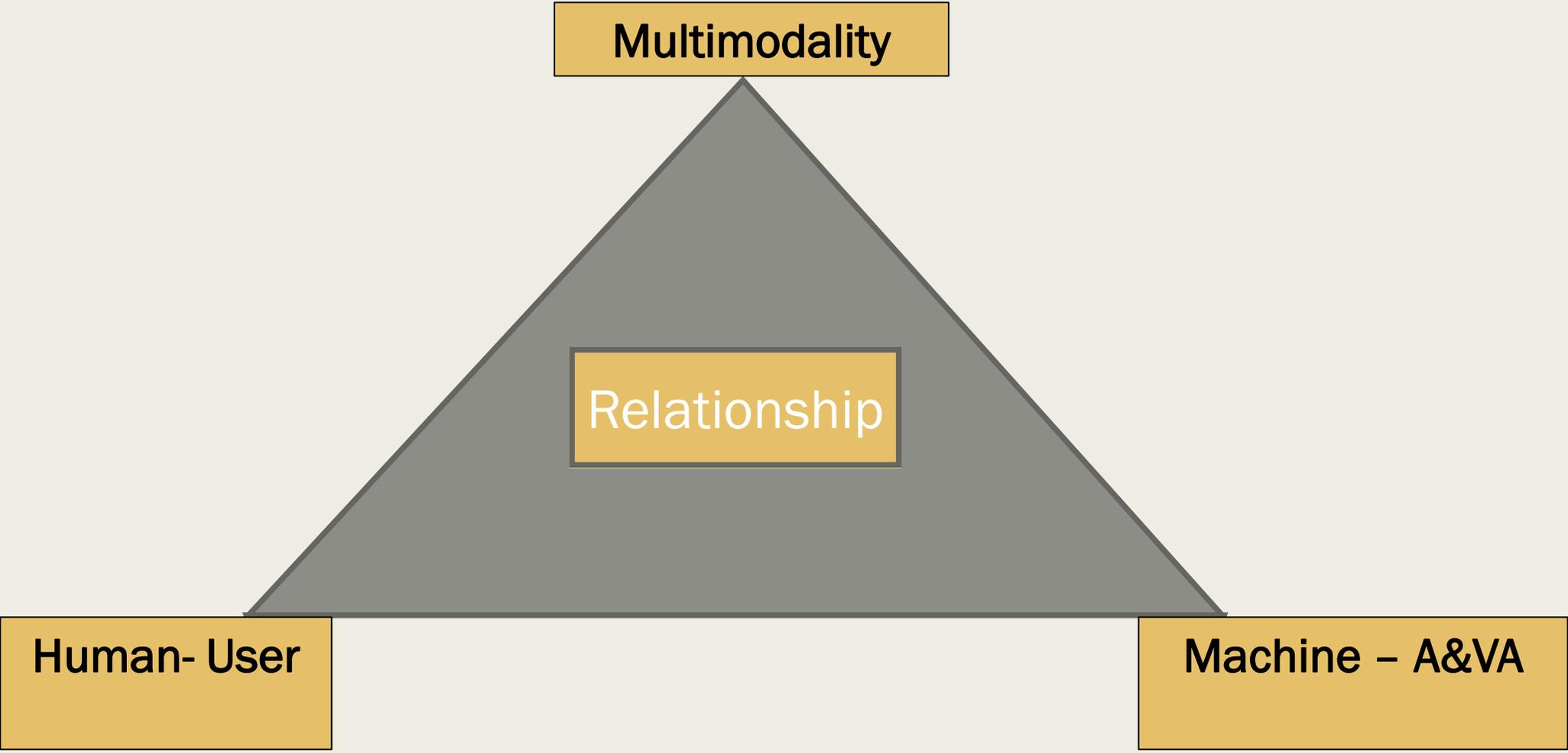
Biometric



Gesture



Cloud



Basic Assumptions about HMI:

- Human Machine relationship is taking place in the interaction and the aspiration is to achieve a dialog that enables task completion whatever the user task is.
- **“Collaboration”** is a process in which two or more participants coordinate their actions towards achieving shared goals” (*Rich et al.*). They introduce the concept of an agent independent of the application that communicates with the user in a three-way communication (agent, app, and user).
- The machine has been personified and **“Collaboration”** is now taking place between two users: one virtual and one human.
- Directly relate them to the concept of **“team work”** as described by *Goldman & Degani (2012)*
- **“Team work”** means being responsive to each other’s needs and mutually supportive to succeed in the joint plan in our case between the Avatars & Virtual Agents and the human end user

Relationship Classification will be the basis for making a decision about which multimodal technologies or interaction types to choose:

We want this We propose 3 main relationships types between human users and Avatars & Virtual Agents:

- Type 1- The Avatar as virtual me
- Type 2 – A personalized/specialized agent
- Type 3 - Me and a random conversational agent

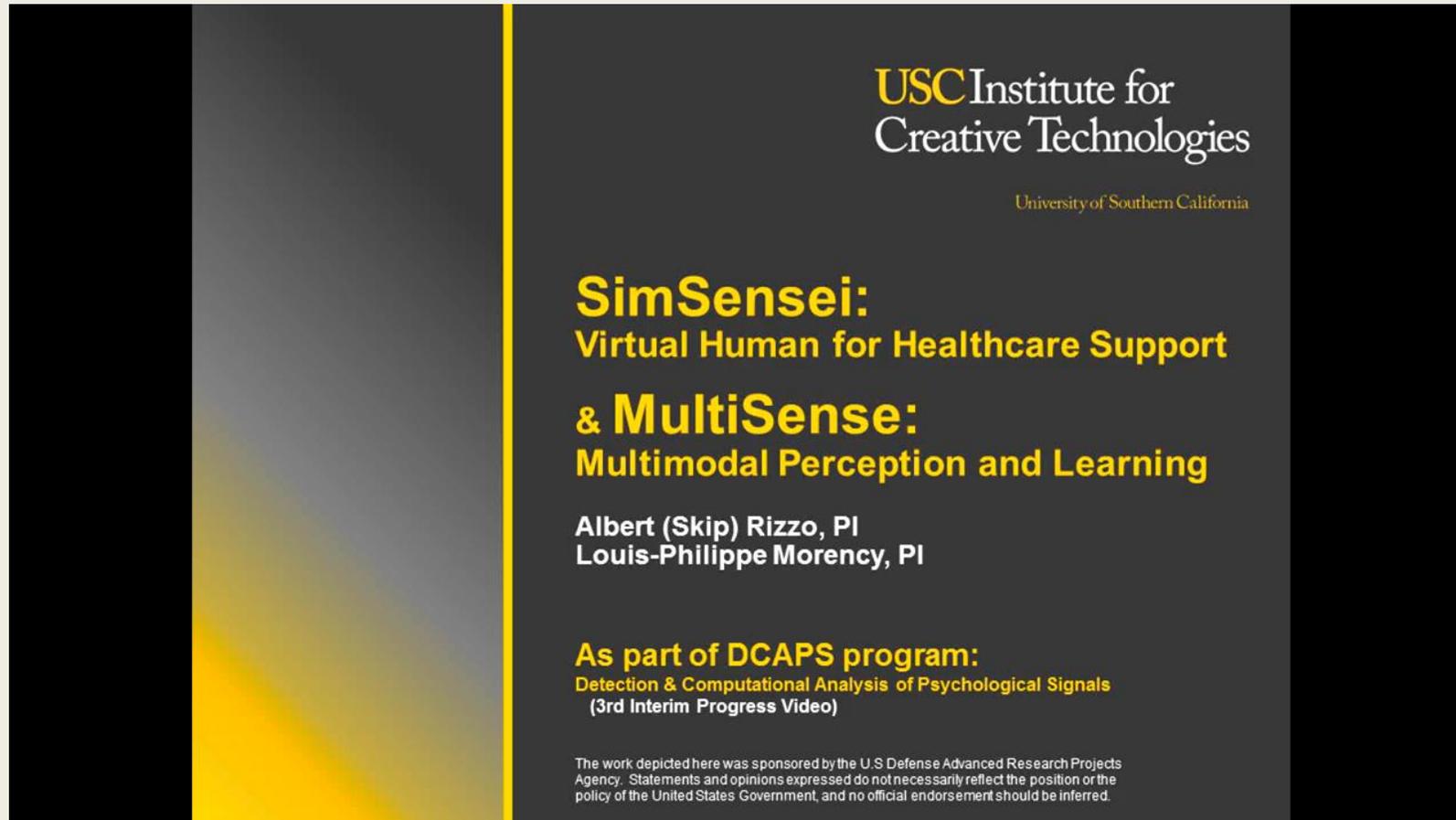
Type 1- The Avatar as virtual me

- Mirror image of the user- an extension
- A persona used to function as an extension for various purposes
- Relationship is self-contained; personal activities are tracked without the need for external intervention.
- Highly personalized according to the profile of the user and his preferences, with the ability to learn patterns and behaviors and the relationship
- Can take on a high degree of intimacy with time
- The preferred modalities will be inherent to the user and his profile and will not change very much.

Type 2- a personalized/specialized Agent

- Specialized Virtual Agent that specializes in a certain type of activities such as: banking and financial services, HMO health agents', education, and government assistance.
- These virtual agents represent a company, an organization or a service provider
- They are not random and have prior knowledge of the user and contact their customer, on a routine basis. Knowledge of profile history and habits.
- type 2 may be as intimate as the personal avatar but in their specific line of business they can make predictions and next best offers.
- They can also connect generally available data with personal data to maximize service.

SimSensei- interactive virtual agent-based clinical decision support tool and



USC Institute for
Creative Technologies
University of Southern California

SimSensei:
Virtual Human for Healthcare Support
& **MultiSense:**
Multimodal Perception and Learning

Albert (Skip) Rizzo, PI
Louis-Philippe Morency, PI

As part of DCAPS program:
Detection & Computational Analysis of Psychological Signals
(3rd Interim Progress Video)

The work depicted here was sponsored by the U.S Defense Advanced Research Projects Agency. Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

Type 3- Me and a conversational agent that is random

- The user and a Virtual Agent with whom he is not familiar, a random interaction with a virtual representative who contacts user while he surfs the internet an online segmentation and targeted advertisement campaign.
- Essentially these random virtual agents can be defined as virtual personas answering real-time questions asked on a website, at click speed, without pauses, 24 hours a day, completely automatically. These VAs can engage in a dialog and help customers make a decision
- User is mostly in a passive and less cooperative mode.
- For design purposes these agents must be very interesting, articulate and very engaging to get user attention and response.

Type 3 example

Samaine Project: Obadiah



Samaine Project: Spike



Multimodal Choices

So how can we design a satisfactory interface for A& VA using multimodal capabilities?

The choice of MMI will be based on the type of relationships (type 1-2-3)

From an interaction point of view -3 factors must be analyzed and modeled before designing a solution:

User

Purpose/task

Environment

The User:

Identifying the target user is a key factor. User preferences for input & output technologies.

The Purpose/Task:

The task the user is trying to complete using the Avatar: Is the Avatar being used to get information, for Q&A, for education, diagnostics or to provide data to someone else. We can differentiate among three types of use cases: active, passive or mixed use.

The Environment

The interaction modality is also influenced by the usage environment.

Users do not usually use speech in a very noisy environment or eye tracking when it's dark outside.

Environment is not only the surrounding and ambience. It also includes the platform on which the virtual agent is used.

Conclusion & Future research

- The area of A&VA is highly connected to MMI
- Yet we do not have enough data regarding the different types of interactions
- No clear methodologically to measure the success of such interaction in this environment.
- In the process of working on usability matrix for A&VA in a multimodal environment

Thanks !

Dr. Nava Shaked

nava@navasaked.com