Applied systems: Practical speech user interface

1th June, 2016

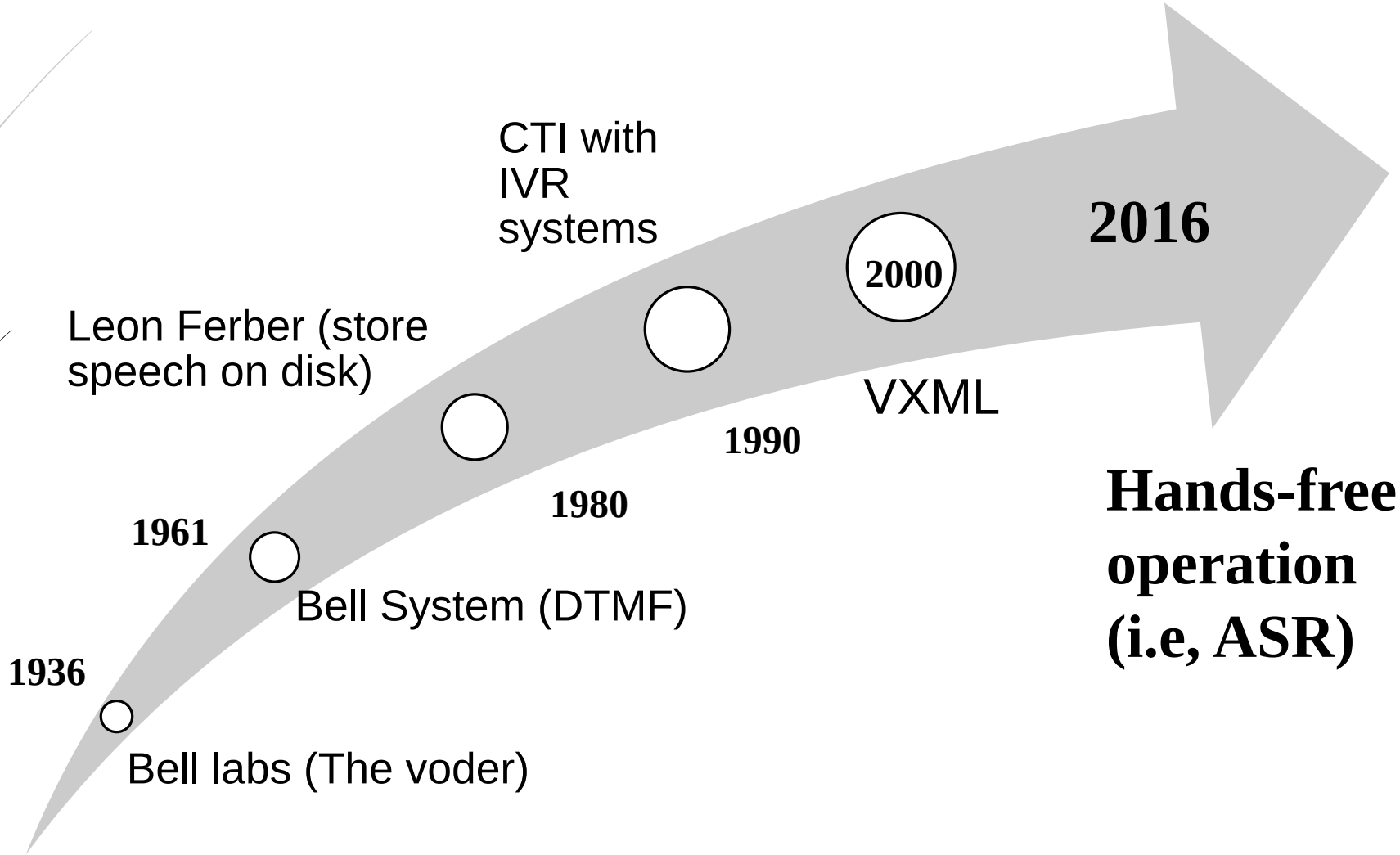
CONTENT

1.1 Introduction

- Frustration by an automated attendant as an IVR (interactive voice response).
- The importance to develop speech user interface (SUI) has gained more acceptance.
- Thus to improve the spoken dialogue is yet a currently challenge.

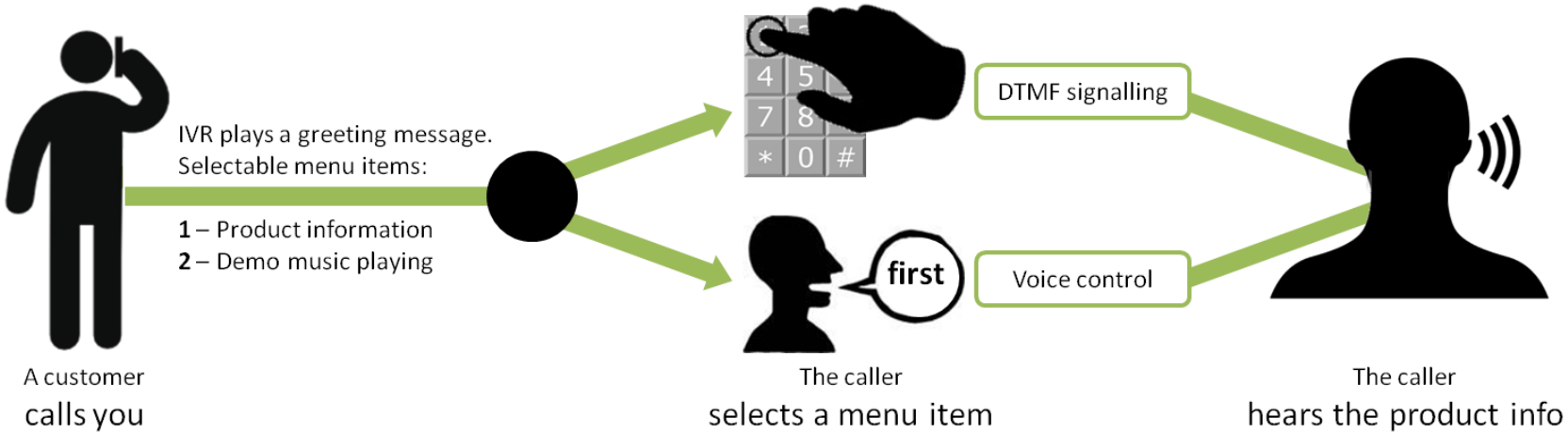
[a]

1.2 Motivation



2 Self-Service Technologies (SSTs)

- IVR as majority self-service
- IVR might simply route calls to skill groups in a call center
- IVR might allow callers to perform self-service
- Most of IVR do a combination of routing and self-service. [1]



[b]

3 The importance of SUI

Drivers of dissatisfactions with SSTs

Category	Description	%Total
Technology failure	The technology simply didn't work, resulting in problems that would affect any user- e.g., broken ATM, Web site refusing to allow login	43
Poor design	Problem with design that would affect some, but not necessarily all users- 17% poor design of technology for usability (e.g., online purchase delivered to cardholder address rather than address of intended recipient of gift).	36
Process failure	Problem with process after successful completion of the initial customer-technology interaction-e.g., problem with billing or delivery	17
Customer-driven failure	Problem in which customers believe they bear some responsibility for the failure-e.g., forgot PIN or password	4

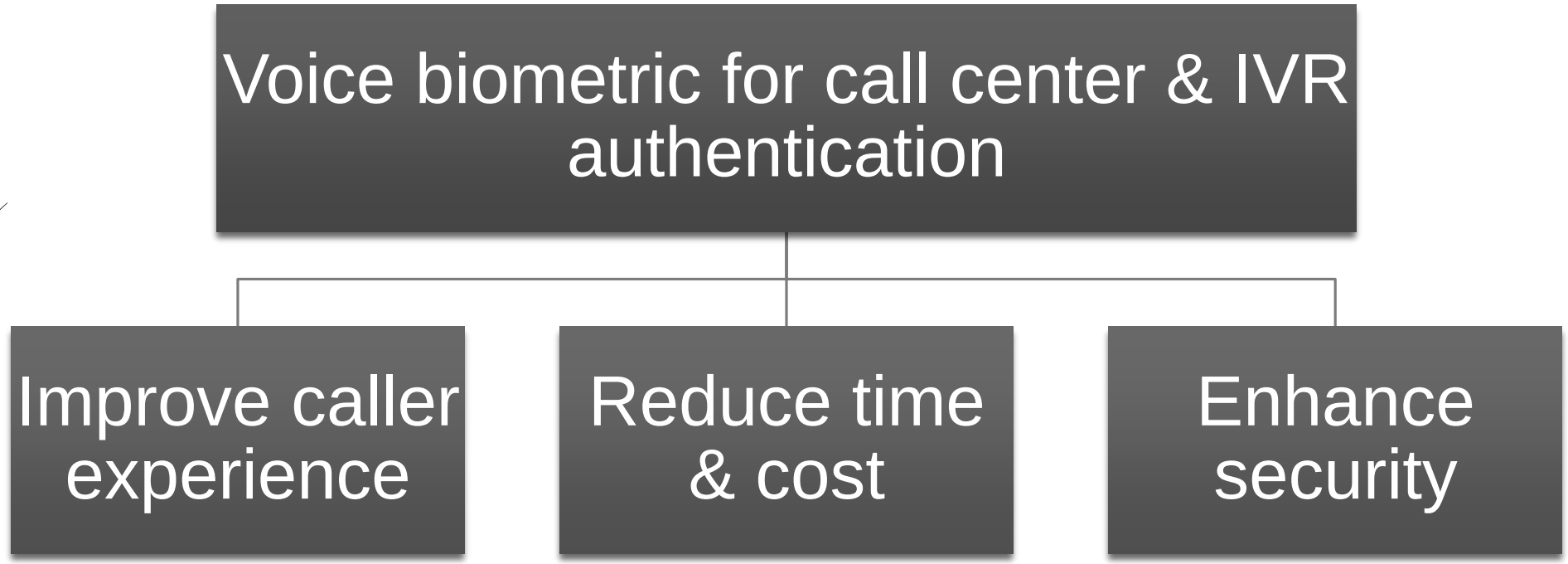
[1]

4 Advances in applied dialog/IVR technology

- 2013 Barclays Wealth used the speaker recognition to verify identity of telephone customers within 30 seconds of normal conversation.
- Two years ago GoVivace Inc. deployed a speaker identification system that allowed search for an individual among millions of speakers .
- Three months ago HSBC announced that they would offer 15 million customers its biometric banking software .

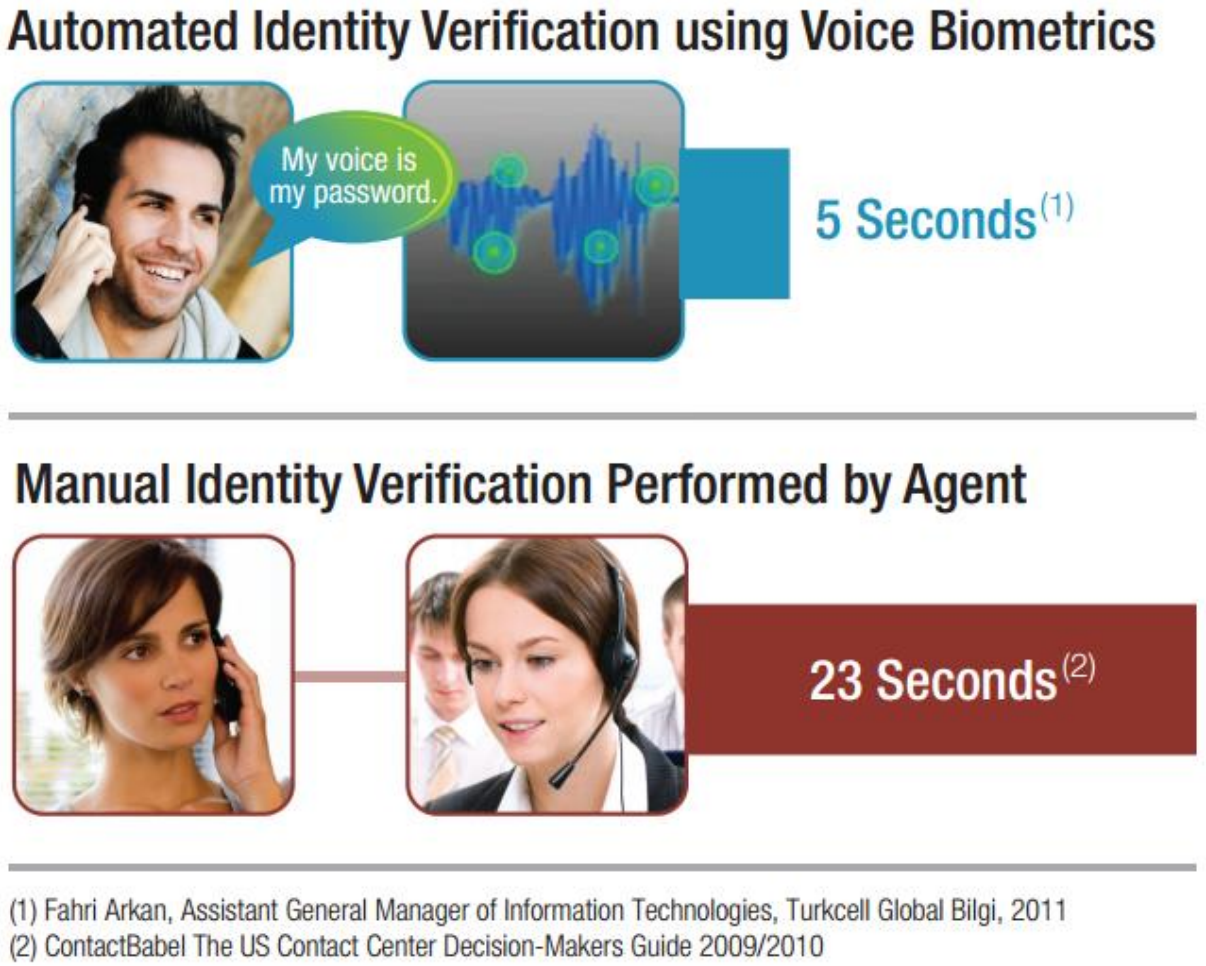
[2]

4 Advances in applied dialog/IVR technology



[3]

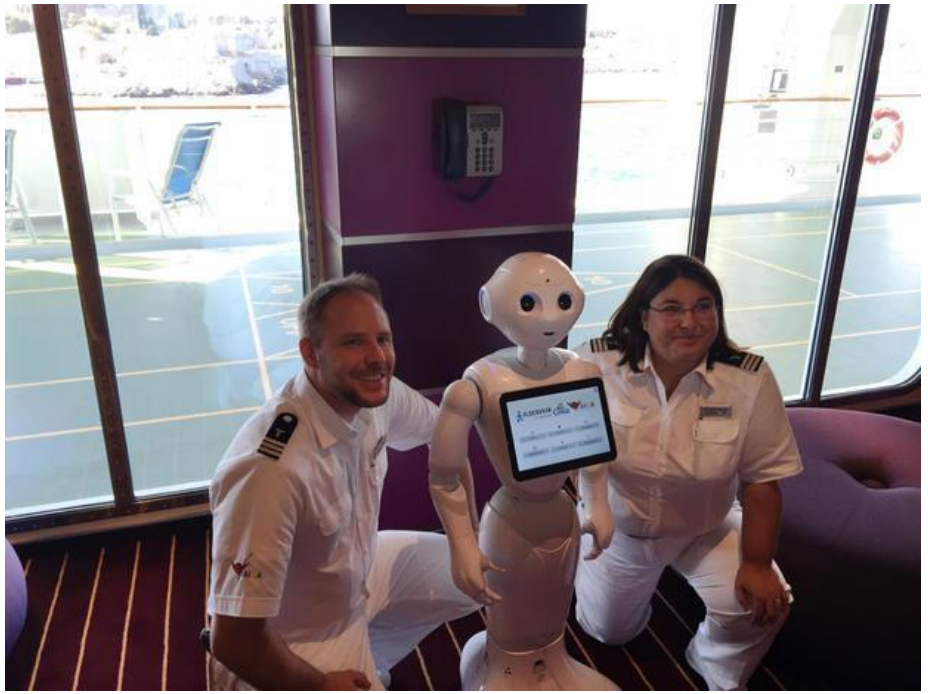
4 Advances in applied dialog/IVR technology



[c]

4 Advances in applied dialog/IVR technology

The Costa Group, an Italian cruise company has got a new member crew: “Pepper”. [4]



[d]

5 Error handling in spoken dialogue

The following example illustrates some speech recognition errors that occurred in a spoken language dialogue between a human user and an automated system:

- **S1:** What city are you leaving from?
- **U1:** Urbana Champaign [OK IN THAT SAME PAY]
- **S2:** Sorry I‘m not sure I understood what you said. What city are you leaving from?
- **U2:** Urbana Champaign [FOR MINUTE SINCE HEY]
- **S3:** I‘m still having trouble understanding you. To go on, I need you to answer the following question. What city are you leaving from?
- **U3:** Chicago [CHICAGO]

(continues)

5 Error handling in spoken dialogue (continuation)

- **S4:** Traveling from Chicago. Where would you like to go?
- **U4:** Huntsville [SEOUL]
- **S5:** Traveling to Seoul. What day did you need to travel?
- **U5:** No, no, I‘m traveling to Birmingham [THE TRAVELING TO BERLIN PM]
- **S6:** Traveling in the afternoon. Okay what day would you be departing Chicago?
- **U6:** Uh, the tenth of august [AT THE TENTH OF AUGUST]
- **S7:** A flight tomorrow I have a Korean Airlines flight departing Chicago at 1:40 PM., arrives Seoul at 5 PM., the next day. Do you want to take that?
- **U7:** My destination is Birmingham [FLIGHT DESTINATION MR WEEKEND]
- **S8:** Traveling on Saturday, August 12th. I have a flight departing Chicago at 1:40 PM., arrives Seoul at…

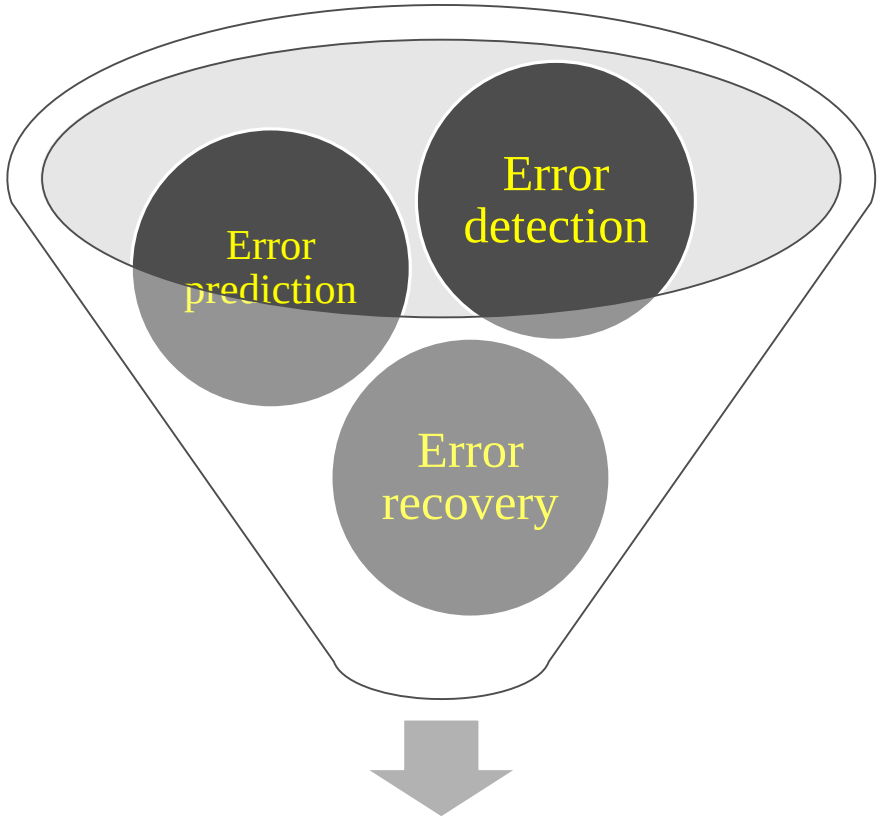
[5]

5 Error handling in spoken dialogue

Try to prevent the errors through careful design of:

- Prompts
- Recognition grammars
- Dialogue flows
- Verification strategies [5]

5 Error handling in spoken dialogue



Processes of error handling

[5]

5.1.1 Error handling in spoken dialogue: *Early error detection*

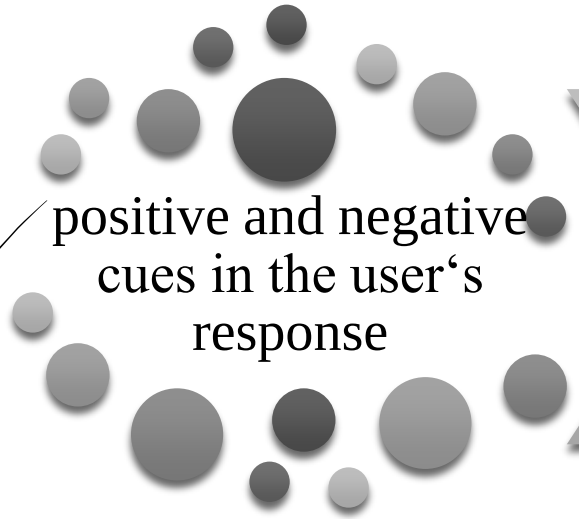
- Using secondary properties of the decoding process
- Comparison of prosodic cues in correctly and incorrectly recognized utterances to predict speech recognition performance.
- The use of combination of cues such as parsing confidence, degree of context shift, and salience to reliably predict errors.
- The use of the concept confidence scores derived from speech recognition confidence scores using a discourse model of what has been said in the dialogue and what entities have been referred to.

5.1.2 Error handling in spoken dialogue: *Late error detection*

Example:

- **S1:** Good morning, this is the talking computer. From which station do you wish to travel?
- **U1:** From Eindhoven to Reuver.
- **S2:** On what day do you want to travel from Eindhoven to Winsum?
- **U2:** I want to travel at 8 o'clock from Eindhoven to **Reuver**
- **S3:** On what day do you want to travel from Eindhoven to Utrecht CS at 8 o'clock in the morning?
- **U3: No, in the evening, at 8 o'clock**. I want to travel from Eindhoven to **Reuver**.

5.1.2 Error handling in spoken dialogue: *Late error detection*



positive and negative cues in the user‘s response

Prosodic information in the user’s utterance

Boundary tone? High pitch? Relatively long delay?

5.2 Error handling in spoken dialogue: *Error prediction*

- Speech recognition features (confidence scores, utterance duration and number of recognized words)
- Natural language understanding features (parsing confidence and context shift)
- Dialog features (prompts and confirmations)

Results: with accuracy rates ranging from 72% to 87%, depending on how much of the dialogue had been seen so far.

[d]

5.3 Error handling in spoken dialogue: *Error recovery*

Example:

- **S1:** How may I help you?
- **U1:** I need to find a train going to New York City [DC I DON'T CARE ON THURSDAY. -5.293949].
- **S2:** Which city do you want to leave from?

(Initial setting is „no confirmation". System asks the next question despite de misrecognition---score exceeds of -4)

- **U2:** Baltimore
- **S3:** Do you want me to find the trains from Baltimore to DC on Thursday at anytime now?

(continues)

5.3 Error handling in spoken dialogue: *Error recovery*

(continuation)

- **U3:** No. I need to leave on Sunday. [DC BETWEEN 1 NIGHT AND 8. -4.916628]
- **S4:** I heard you say go to DC between 1 night and 8. Do you want me to find you the trains from Baltimore to DC on Thursday between 1 night and now?

(System adapts is strategy: there has been at least one misrecognition within the previous four utterances. System uses implicit confirmation „I heard you say…“ combined with mixed initiative.

5.3 Error handling in spoken dialogue: *Error recovery*

Example of how confidence scores may be associated with grounding actions:

Confidence level	Action
High	Accept
Fairly high	Display understanding (implicit verification)
Fairly low	Clarification request (explicit verification)
Low	Reject

5.4 Some new approaches to error handling

Error handling based on the strategies used by humans in dialogue systems:

- S1: Do you see a wooden house in front of you?
- U1: I pass the wooden house now [YES CROSSING ADDRESS NOW]: Nonunderstanding
- S2: Can you see a restaurant sign?

- U1: I can see a blue building
- S1: Blue
- U2: No
- U3: Brown

5.4 Some new approaches to error handling

Error handling in the RavenClaw System:

R: indicates the recognition result

P: its semantic representation

- S1: Where are leaving from?
- U1: the airport

  R: LIBERTY and WOOD

  P: [departure= LIBERTY and WOOD]
- S2: Leaving from Liberty and Wood? Is that correct?
- U2: Nope

  R: NO

  P: [NO]

5.4 Some new approaches to error handling

Error handling in the DIHANA Project – A Corpus-Based approach:

To Bilbao on March second. When does it leave and arrive? An how much is it? Some possible transitions to system sates are as follows:

Confirm Destination

Confirm Departure-Date

Give Departure-Hour

Give Arrival-Hour

Give Price

5.4 Some new approaches to error handling-**DIHANA Project**

High scores for destination and date but a low score for price:

(**U:** Question: Depart_hour, Arrival_hour)>(**S:** Answer: Depart_hour, Arrival_hour)

It leaves at six o'clock and arrives at eight o'clock

Given a user turn consisting of three input frames with confidence scores for the syntactic and semantic analyses in parentheses:

(AFFIRMATION) * [0.78, 0.83]

DESTINATION: Bilbao * [0.82, 0.85]

TRAIN-TYPE: Talgo * [0.62, 0.66]

(DEPART-HOUR) * [0.91, 0.95]

(PRICE) * [0.55, 0.46]

5.4 Some new approaches to error handling-**DIHANA Project**

A frame consisting of the following dialogue acts is produced:

(U: Affirmation: Destination, Train_type)

(U: Question: Depart_hour)

(U: Question:Price)

Using semantic generalization, the following alternative frames are produced for this input:

3 (U: Question: Depart_hour) (U: Question:Price)

4 (U: Affirmation: Destination, Train_type) (U: Question: Depart_hour)

5 (U: Question: Depart_hour, Price)

6 (U: Affirmation: Destination, Train_type) (U: Question:Price)

7 (U: Affirmation: Destination) (U: Question: Depart_hour)

5.4 Some new approaches to error handling-**DIHANA Project**

In the first example, all three attributes have high confidences scores, so no confirmation is given:

U0: Hello, I want to go from Burgos to Bilbao on March the second

U0: (DEPART_HOUR) * [0.97, 0.92]

ORIGIN: Burgos * [0.97, 0.92]

DESTINATION: Bilbao * [0.97, 0.92]

S0: The train leaves at half past three.

5.4 Some new approaches to error handling-**DIHANA Project**

In the next example, the departure date has a low confidence score, so this attribute is confirmed:

U0: Hello, I want to go from Burgos to Bilbao on March the second

U0: (DEPART_HOUR) * [0.97, 0.92]

ORIGIN: Burgos * [0.97, 0.92]

DESTINATION: Bilbao * [0.87, 0.82]

DEPART_DATE: 02-03-2004 * [0.36, 0.33]

S0: Do you want to leave on March the second?.

6 Challenges

- The design of IVRs can benefit from an understanding of service science.
- It can be difficult to get the people to switch from full service to self service
- The characteristics of call centers and effective call center agents can inform the design of IVRs.
- Dialog System to old people and/or with speech difficult could be very hard to develop.

6 Conclusions

- We have seen the evolution of spoken dialogue over eighty years and speech recognition, from the speech synthesis “The Voder” to nowadays with ASR and dialogue management.
- The importance the SUI for the IVR systems.
- We have learned, some approaches to error handling for spoken dialogues, furthermore, how to measure the errors and how to avoid them.

THANK YOU

References

1. Lewis, J.R (2011). *Practical speech user interface design.* CRC Press
2. https://en.wikipedia.org/wiki/Speaker_recognition
3. Nuance Communications (2012) *Voice biometric for call center & IVR authentication.* Nuance Communications Inc. http://www.nuance.de/ucmprod/groups/enterprise/@web-enus/documents/collateral/nc_019968.pdf
4. http://www.speechtechblog.com/2016/04/26/speech-robots-take-seas#axzz4A3SsuTFs
5. Jokinen, K., McTear, M. (2010). *Spoken dialogue systems*. University of Helsinki (Jokinen) and University of Ulster (McTear). Morgan & Claypool publishers.

Images sources

a. https://tstotips.files.wordpress.com/2013/05/homer-angry-phone.gif
b. http://www.codeproject.com/KB/cs/746512/how_the_ivr_menu_system_works.png
c. Nuance Communications (2012) *Voice biometric for call center & IVR authentication.* Nuance Communications Inc. http://www.nuance.de/ucmprod/groups/enterprise/@web-enus/documents/collateral/nc_019968.pdf
d. http://www.speechtechblog.com/2016/04/26/speech-robots-take-seas#axzz4A3SsuTFs
e. http://www.gampmedia.com/what-is-reputation-management/