

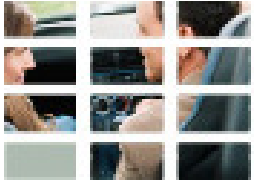
Voice Control becomes Natural

ITU-T FOCUS GROUP CarCom -- SPEECH IN CARS

Dr. Udo Haiber

Torino, Italy, October 16, 2009

SVOX 



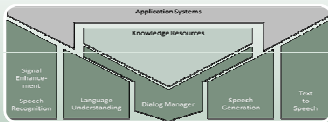
Overview

SVOX 

Company



What is Natural?



Involved Components



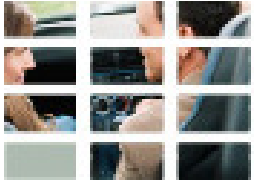
Focus Change



Approach



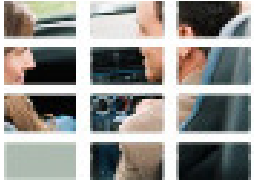
Conclusion



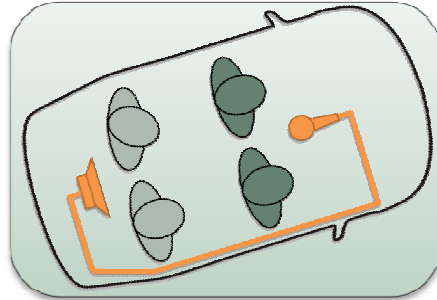
Company - Overview

- Founded in 2000 as university spin-off
- Privately held, headquartered in Zurich, Switzerland, offices in Germany and USA
- More than 100 permanent employees plus more than 100 local language experts
- In January 2009 acquired Speech Processing Group of Siemens AG
- In August 2009 opened office in Ulm, Germany
- World's experts in Speech Technology

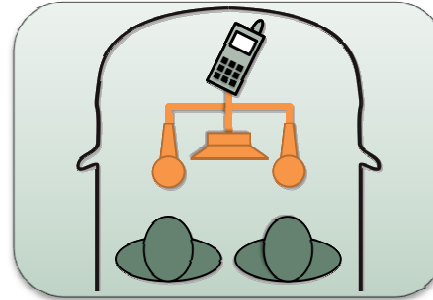




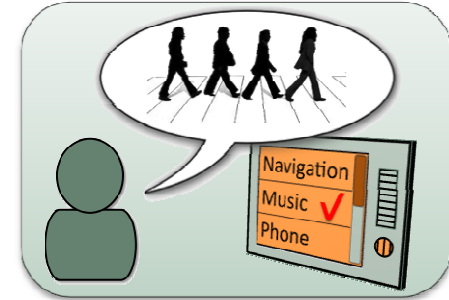
Company – Product portfolio



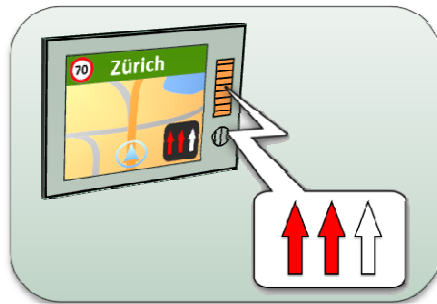
In-Car
Communication



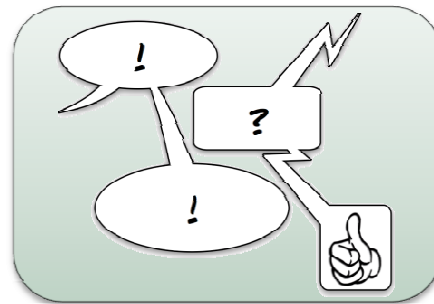
Acoustic Signal
Enhancement



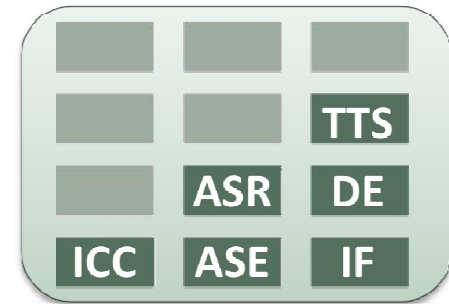
ASR Engines



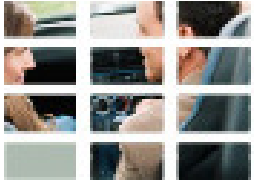
TTS Engines



Dialog Engine



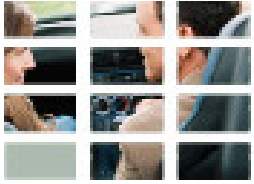
Integration
Framework



Company - Customers & Deployments

The map shows green regions in North America, South America, Europe, and Australia, and grey regions in Africa, Asia, and parts of Europe. Surrounding the map are logos for various companies and automotive brands:

- ASUS**
- BMW**
- Clarion** (CAR AUDIO & VIDEO)
- Continental** (AUTOMOTIVE)
- CHRYSLER**
- EB**
- Ferrari**
- Panasonic**
- AsahiKASEI**
- HARMAN/BECKER** (AUTOMOTIVE SYSTEMS)
- Audi**
- HYUNDAI**
- ANDROID**
- INFINITI**
- acer**
- INTRINSYC**
- xanavi**
- JVC** (The Perfect Experience)
- VW**
- Microsoft Auto**
- SKODA** (AUTO)
- Mercedes-Benz**
- SEAT**
- PEUGEOT**
- NOKIA** (CONNECTING PEOPLE)
- NISSAN**
- NAVIGON**
- Motors**

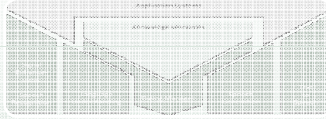


Overview

SVOX  Company



What is Natural?



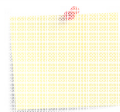
Involved Components



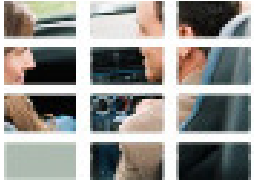
Focus Change



Approach



Conclusion

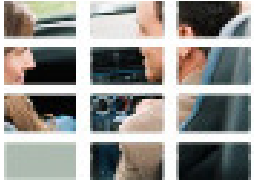


What is Natural – a Paradigm

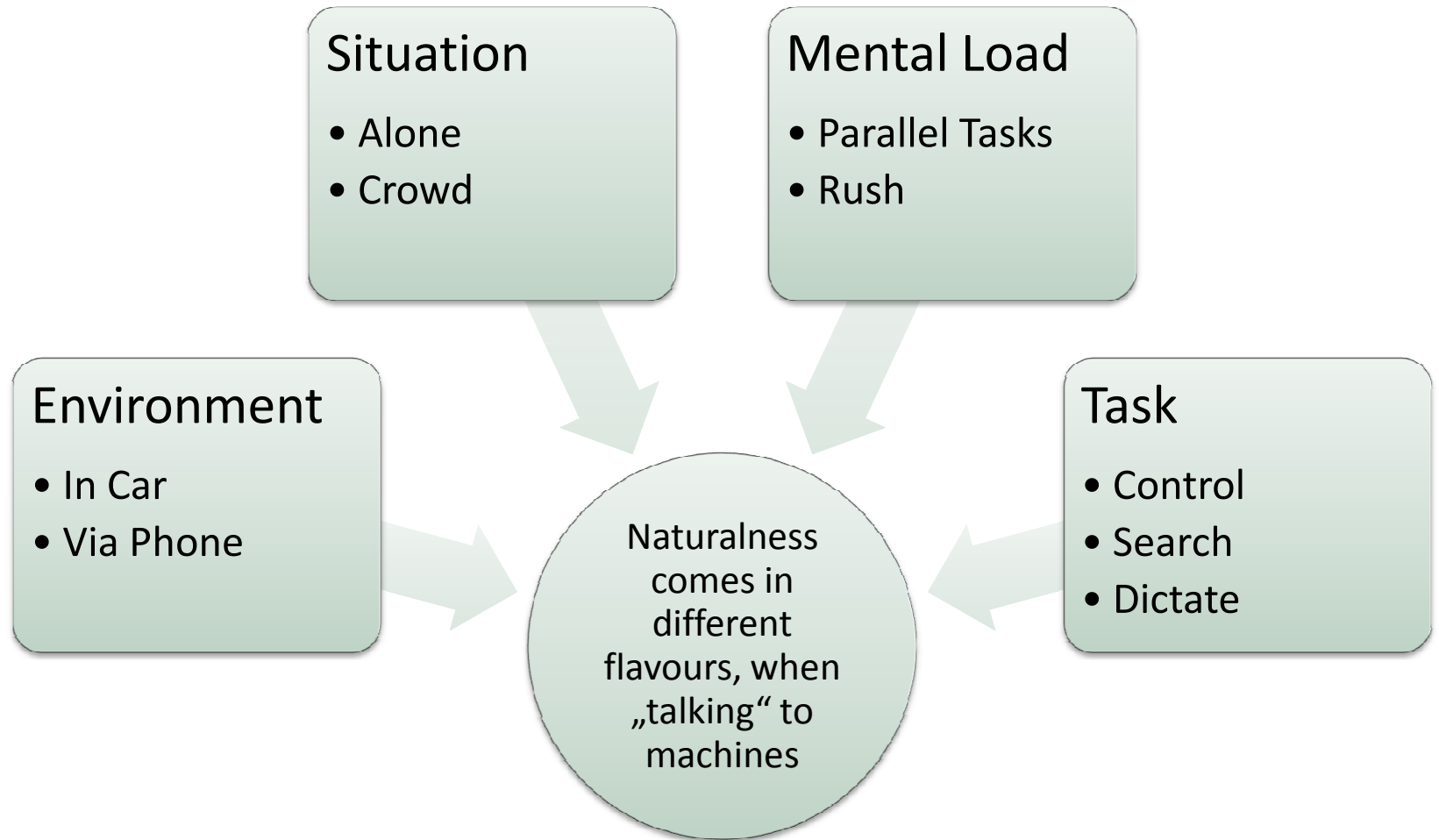


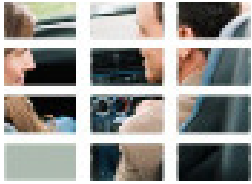
“Reports and theses in speech recognition often begin with a cliché, namely that speech is the most natural way for human beings to communicate with each other”

Melvyn J. Hunt, 1992



What is Natural – with Machines





What is Natural – Task dependency

■ Control of devices

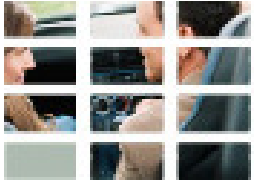
- Short commands, less conversational elements
- Synonyms are important
- Interpretation must be possible (action=„call“, name=„John Smith“, properties=„mobile“)
- „Call John Smith mobile [please]“

■ Search in huge lists

- No commands, only search terms (spiele „spiel das lied vom tod“)
- No conversational elements at all, just conjunction of terms
- No interpretation
- „John“, „John Smith“, „John Smith of SVOX“

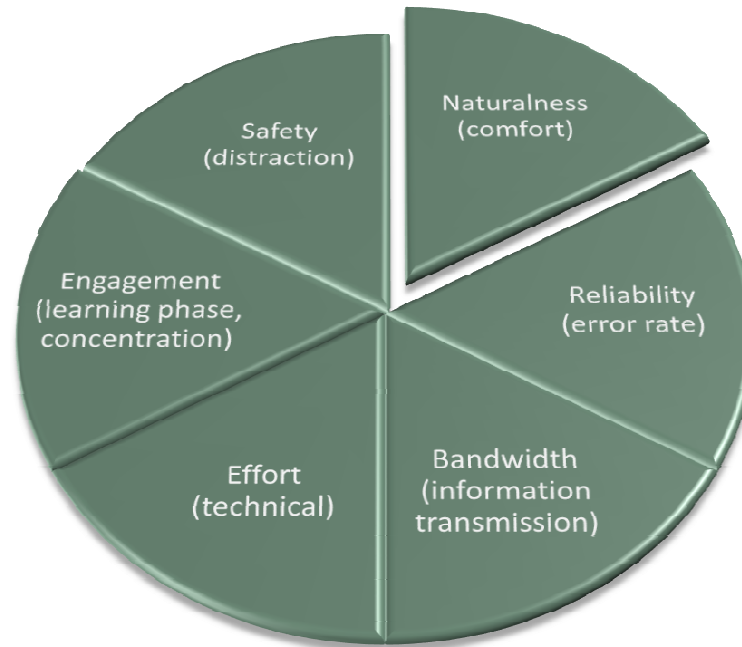
■ Dictate a text

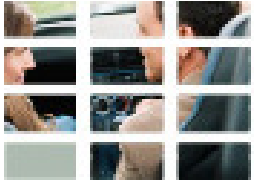
- Commands, conversational elements, Synonyms as required in text
- No interpretation
- „He mentioned, that we should call John Smith. What do you think?“



What is Natural – is that the right question at all?

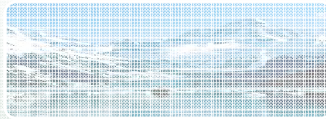
... at least there are more criteria to decide using a speech interface ...



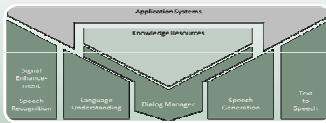


Overview

SVOX  Company



What is Natural?



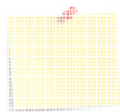
Involved Components



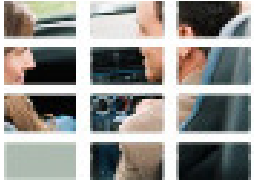
Focus Change



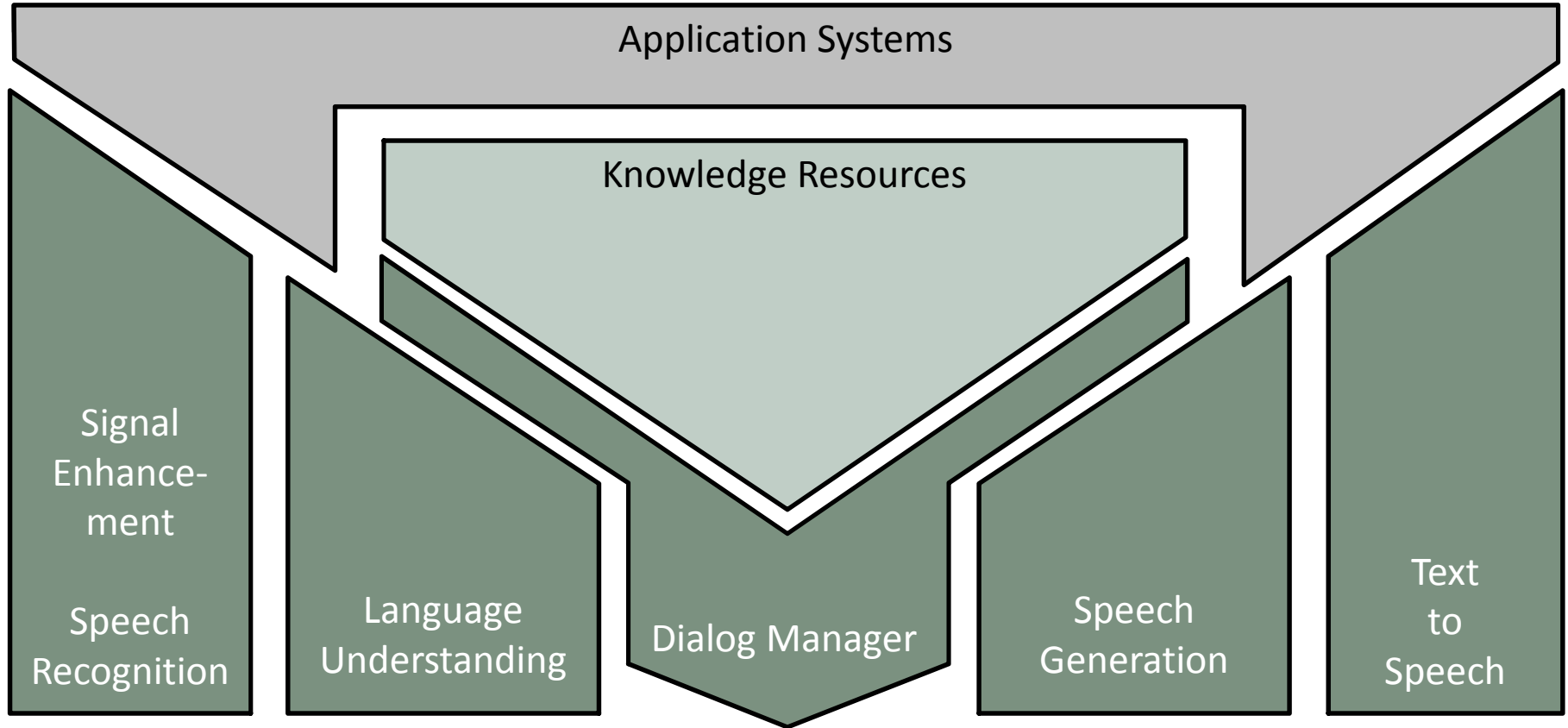
Approach

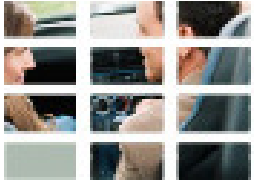


Conclusion



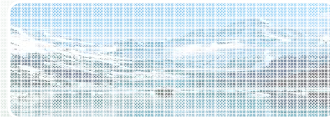
Components of a Speech Dialog System (SDS)



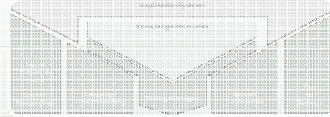


Overview

 SVOX Company



What is Natural?



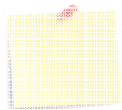
Involved Components



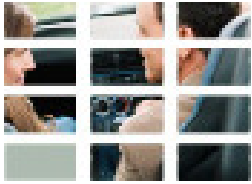
Focus Change



Approach



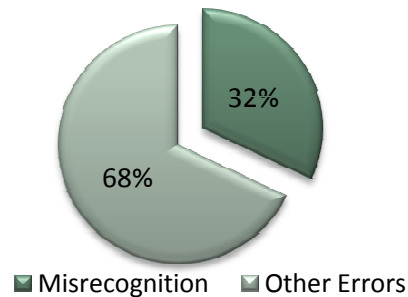
Conclusion



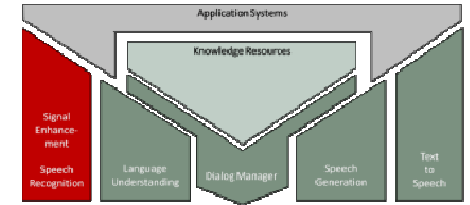
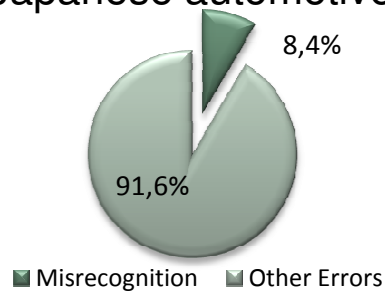
Shift of development focus

Different usability studies lead to the same thesis:
Misrecognition is NOT the major problem

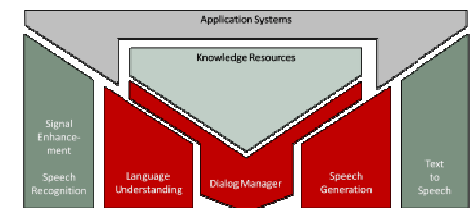
■ European automotive OEM (2005, 50 subjects)

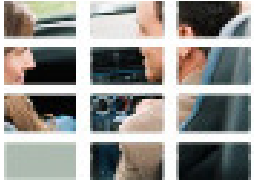


■ Japanese automotive supplier (2009, 245 subjects)



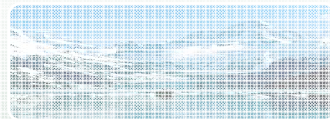
Move focus from
technology to *user*
 centric development



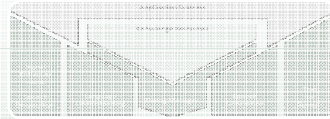


Overview

 Company



What is Natural?



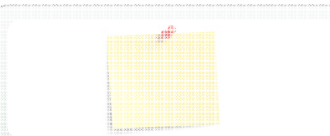
Involved Components



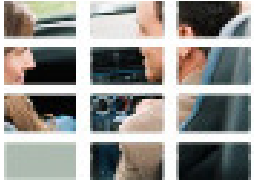
Focus Change



Approach

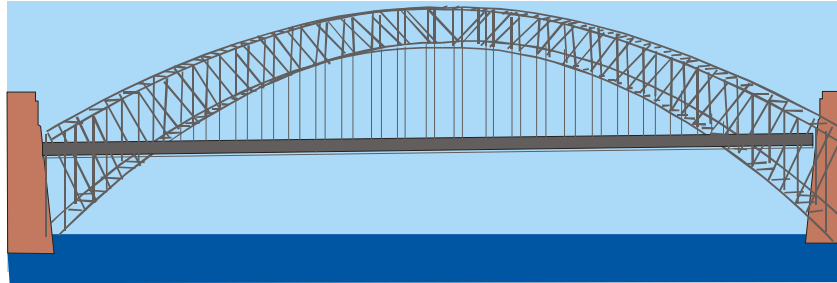


Conclusion



Objective of User centric development

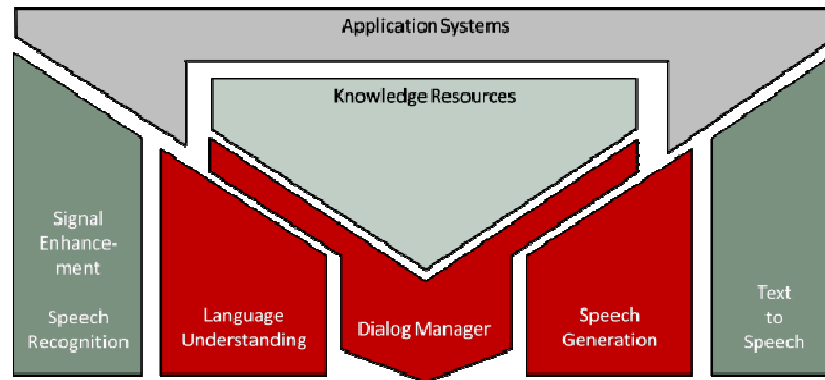
What to do?

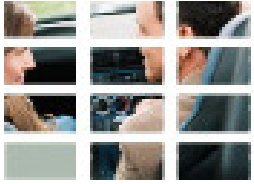


How to do?

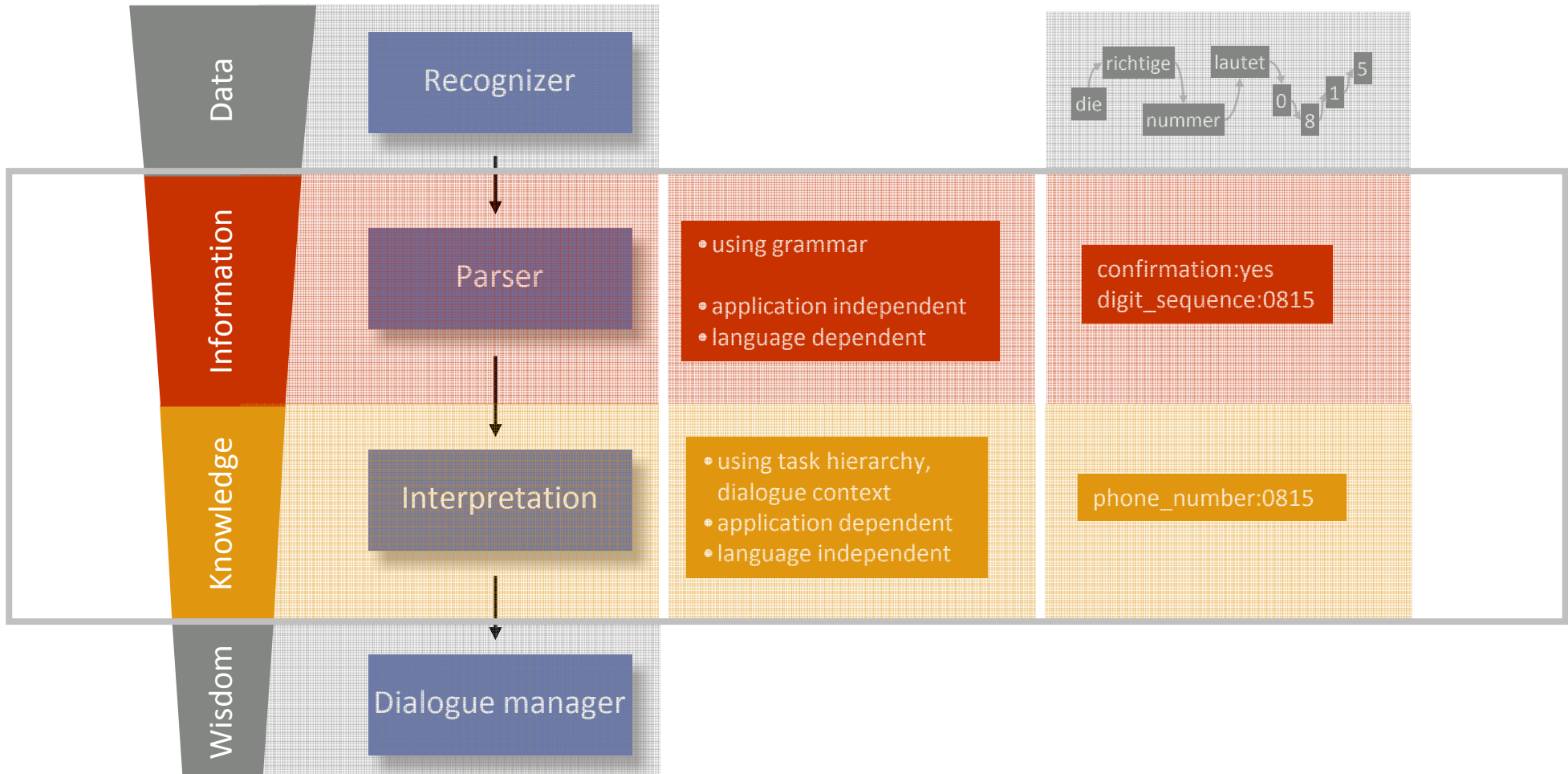
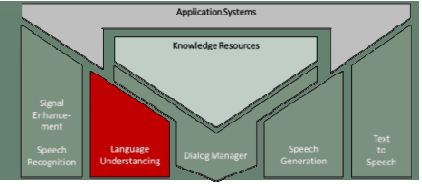


Bridge the gap between “what” and “how” using speech dialogue methods

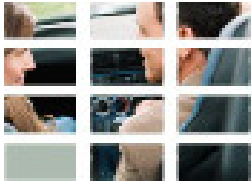




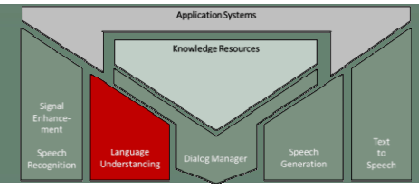
Pyramid of Language Understanding^{*)}



^{*)} Hey, Jonathan (2004). The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link



Challenges of Language Understanding



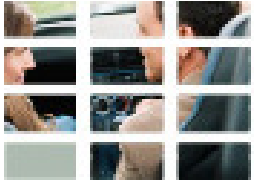
Objective: deliver the meaning of spoken input

■ Ill-formed input

- While speaking, humans are not able to deliver linguistically correct sentences

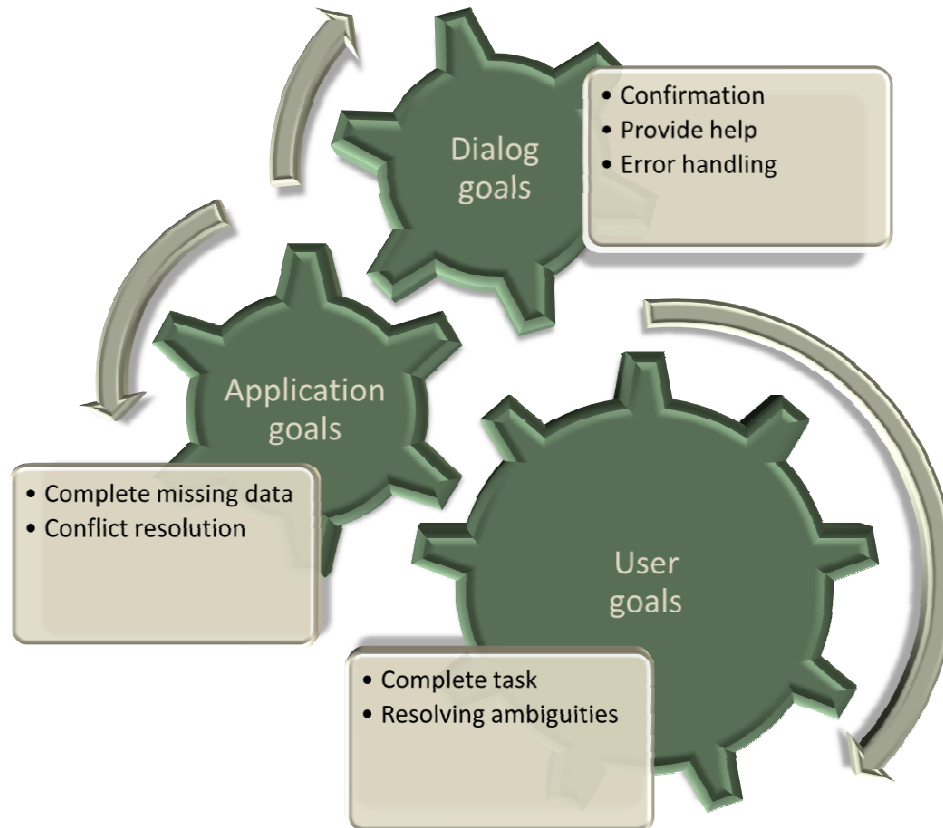
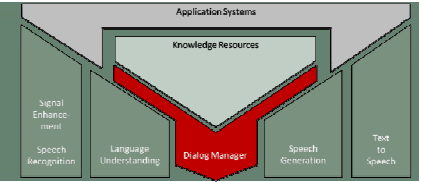
■ Ambiguity in natural language

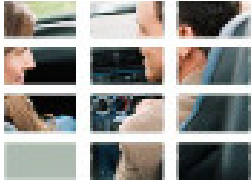
- *Lexical ambiguity*: a word can belong to more than one part of speech.
e.g. book is a noun and a verb
- *Sense ambiguity*: a word can have different meanings
e.g. bank (financial institution vs. side of a river)
- *Structural ambiguity*: relationship between phrases in a sentence is ambiguous
e.g. “John booked (a seat on the train)” vs. “John (booked a seat) on the train”



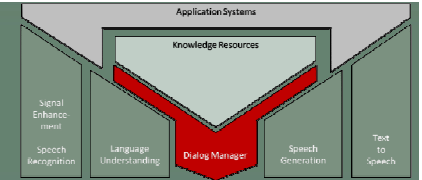
Dialog Management

Decide for next dialog step considering all active goals



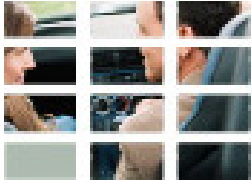


Challenges of Dialog Management

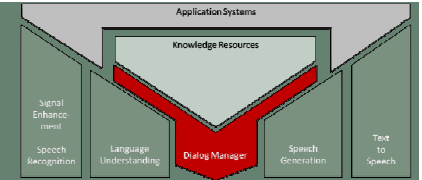


Objective: Minimize the number of steps/interactions

- What must be known by the system to fulfill tasks?
- What could be inferred?
- Which information must come from the user?



Dialog Control

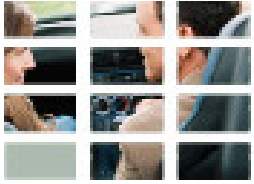


Initiative

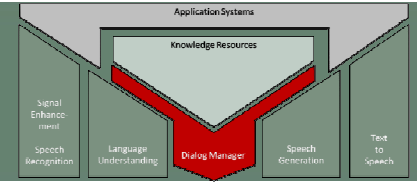
- System driven
- User driven
- Mixed-initiative

Flow

- Finite state models
- Form-based approaches



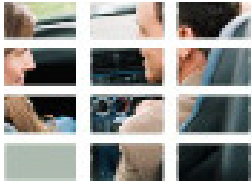
Dialog Control - Initiative



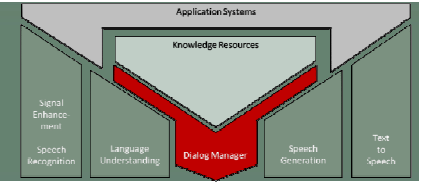
System Initiative

- Shortcuts to allow efficient control
„Call John on his mobile“
 - Mainly used by experts
-
- Guided dialog as help
„Would you like to call somebody?“
 - Mainly used by novice or occasional users

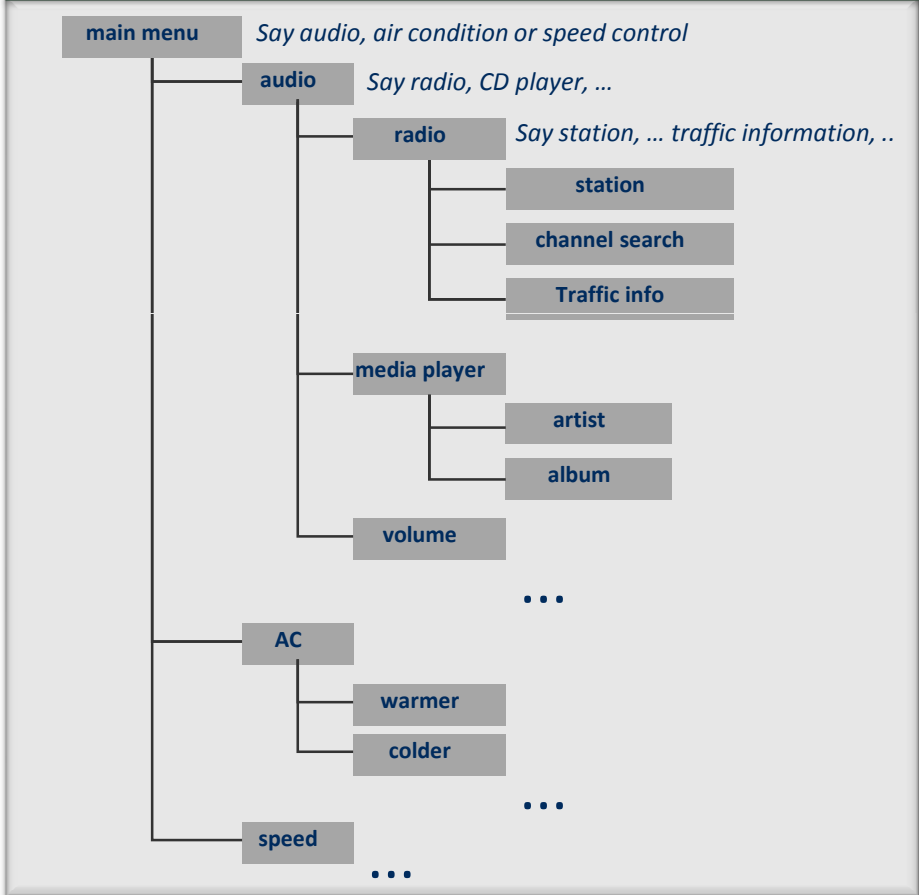
User Initiative



Dialog Control - Flow

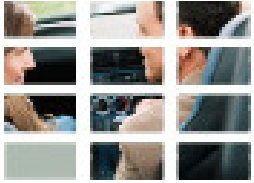


State based systems model the dialog flow



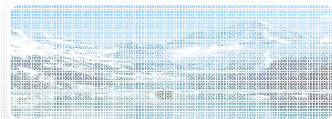
Frame based systems model the dialog content

| Action frame | Slot 1 | Slot 2 |
|----------------------|----------------------|---|
| select radio station | <station> | |
| channel search | forward/ backward | |
| traffic info | on / off | |
| audio | high / low | front / rear |
| air condition | warmer / colder | Left / right / front / rear / everywhere |
| media player | <artist> | <album> |

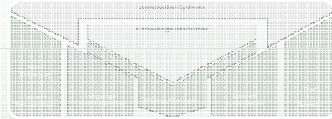


Overview

 Company



What is Natural?



Involved Components



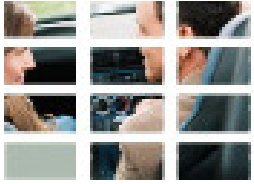
Focus Change



Approach



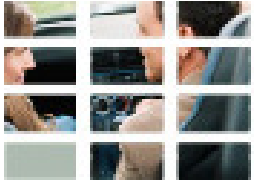
Conclusion



Conclusion

*Do not overcharge speech technology
(deploy only apps, which are really usable)
Expectation must be met across all users
(beginners and experts in one system)*

*Design Voice control with focus on User
(natural dialog apps are easy to implement with a
frame-based engine)*



SVOX – Your Dialog Partner



Contact

SVOX AG
Baslerstrasse 30
CH-8048 Zürich
Switzerland

Phone +41 43 544 06 00
Fax +41 43 544 06 01
info@svox.com
www.svox.com