



Applications of speech-to-text in customer service.

Dr. Joachim Stegmann – Deutsche Telekom AG, Laboratories



Life is for sharing.

Contents.

1. Motivation

2. Scenarios

2.1 Voice box / call-back

2.2 Quality management

3. Technology

4. Results and conclusion



1. Motivation.

Deutsche Telekom Laboratories.



Motivation.

Why do we need innovation in customer service?

Today



Customer not
optimally served



Agent
stressed

Goals

- New options (e.g. audio-visual interaction).
- **Reduce waiting time.**

Increase customer
satisfaction

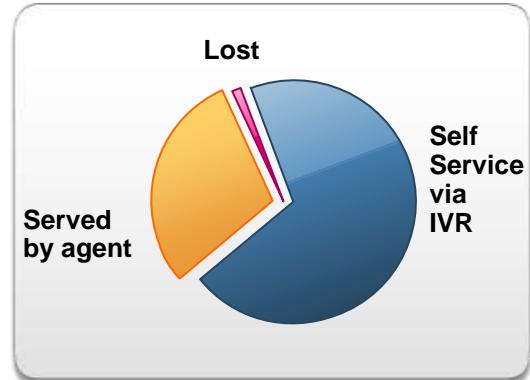
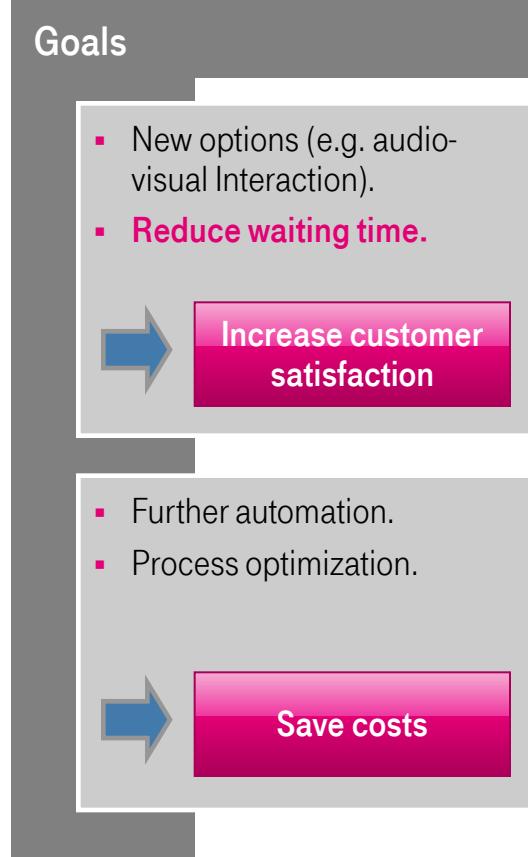
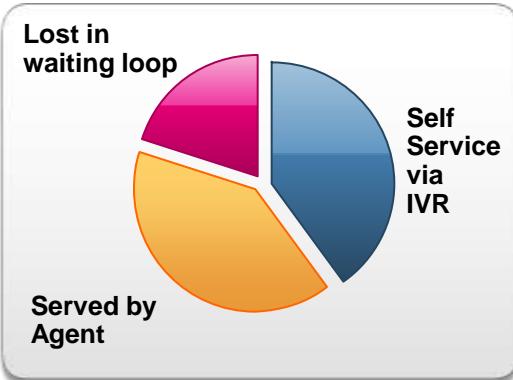
- Further automation.
- Process optimization.

Save costs

Tomorrow



New/ innovative
Solutions



2. Scenarios.

2.1. Voice box / call-back.

Scenario call-back. The idea is not new, ...

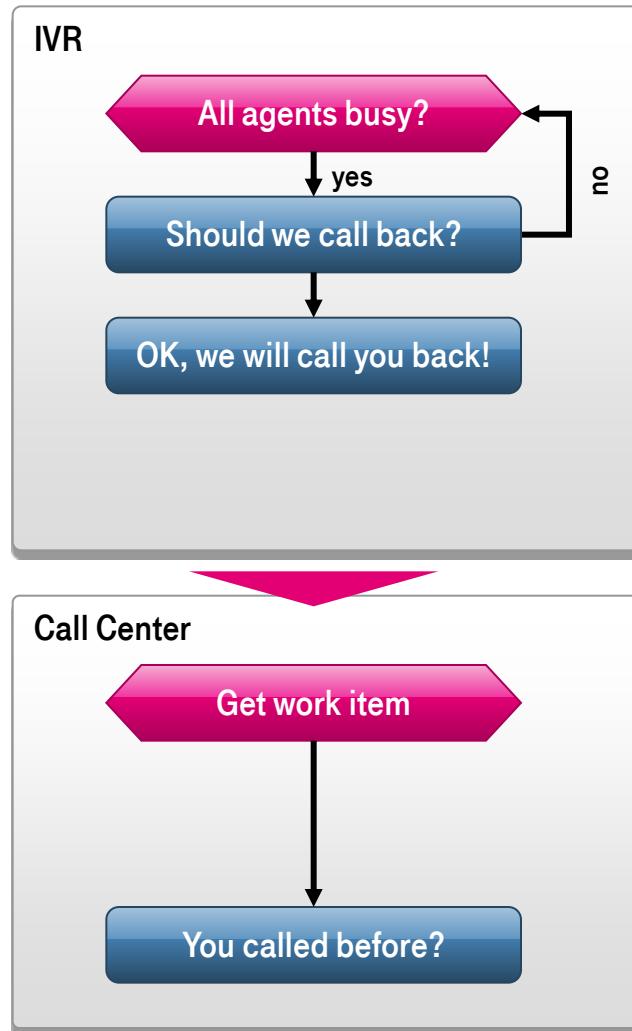
Don't let the customer wait!

Idea

- Customer is called back as soon as an agent is available.
- Already tested.

Problems

- Agent is not aware of the history.
- Customer has to explain problem repeatedly.



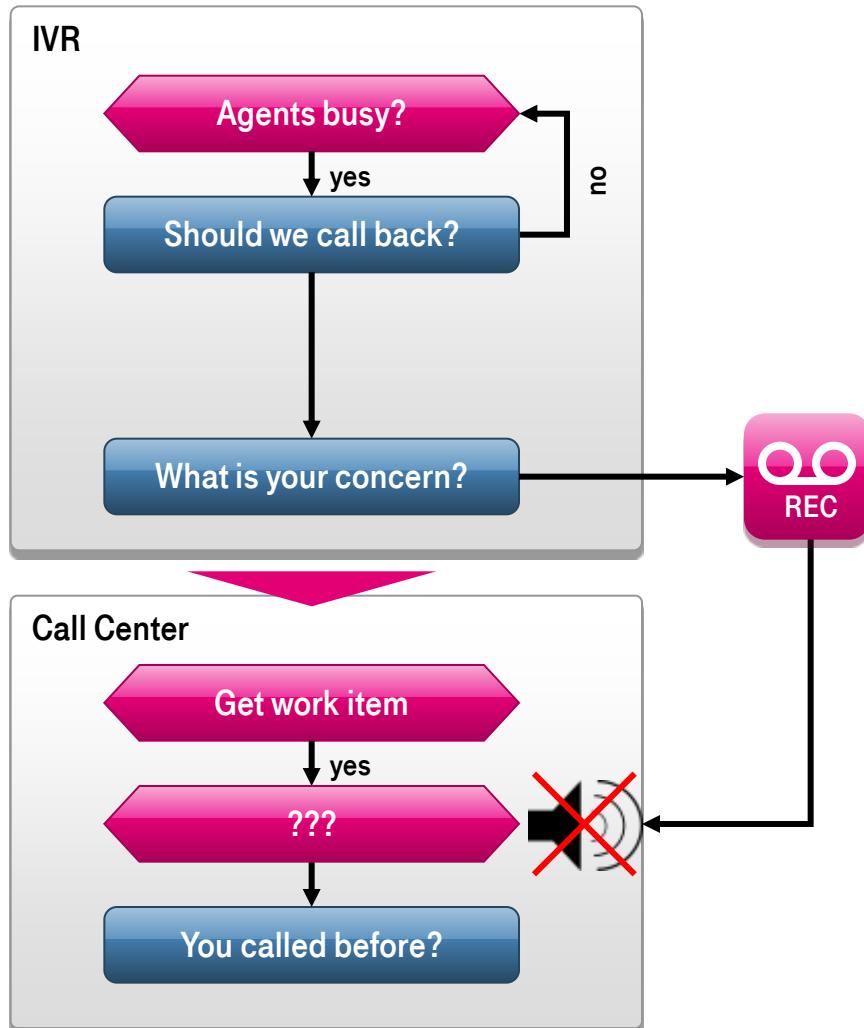
Scenario call-back. ... history would help, ...

Idea

- Record customer concern.
- Customer feels served.

But

- Some agent work places do not allow playing audio files.
- Listening to audio files is time consuming.



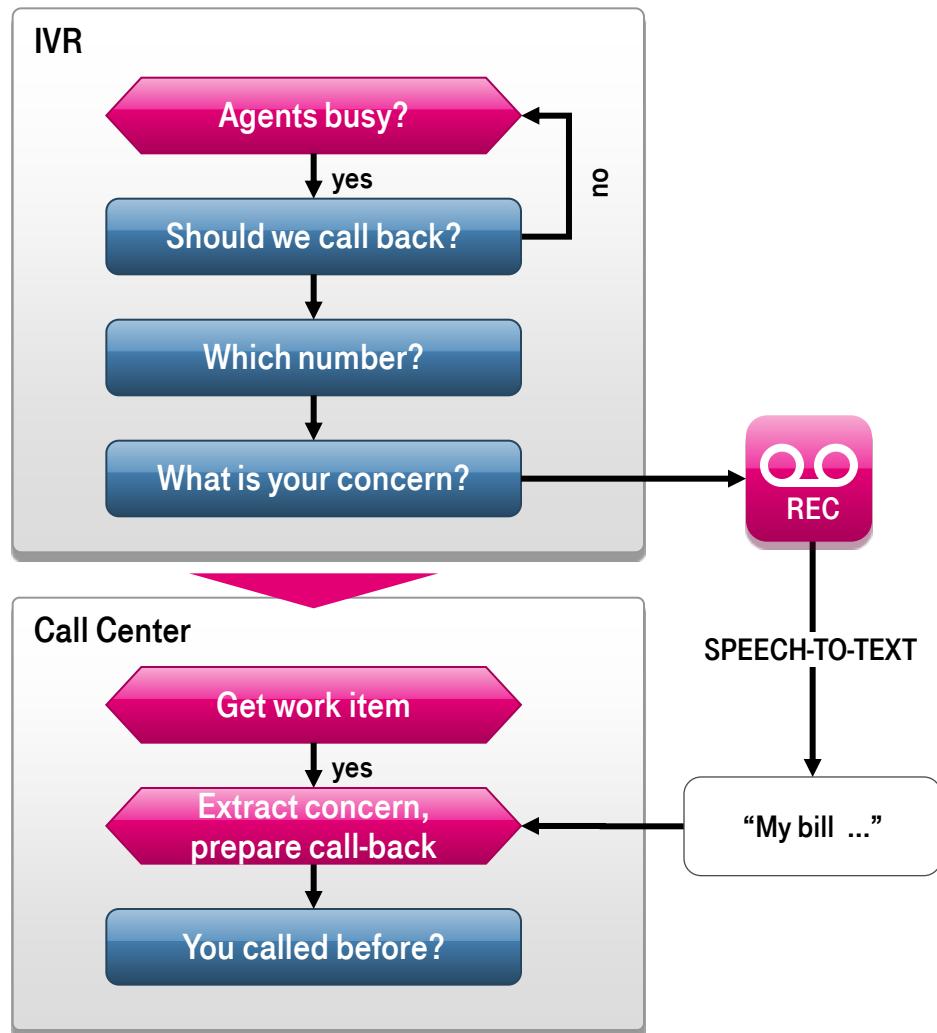
Scenario call-back. ... and is made available via speech-to-text.

Customer

- Is offered a call-back.
- May leave a call-back number.
- Specifies his/her concern.

Agent

- Extracts concern from text.
- Prepares call-back (e.g. by having a look at the customer's bill).
- Calls back.



2. Scenarios.

2.2. Quality management.

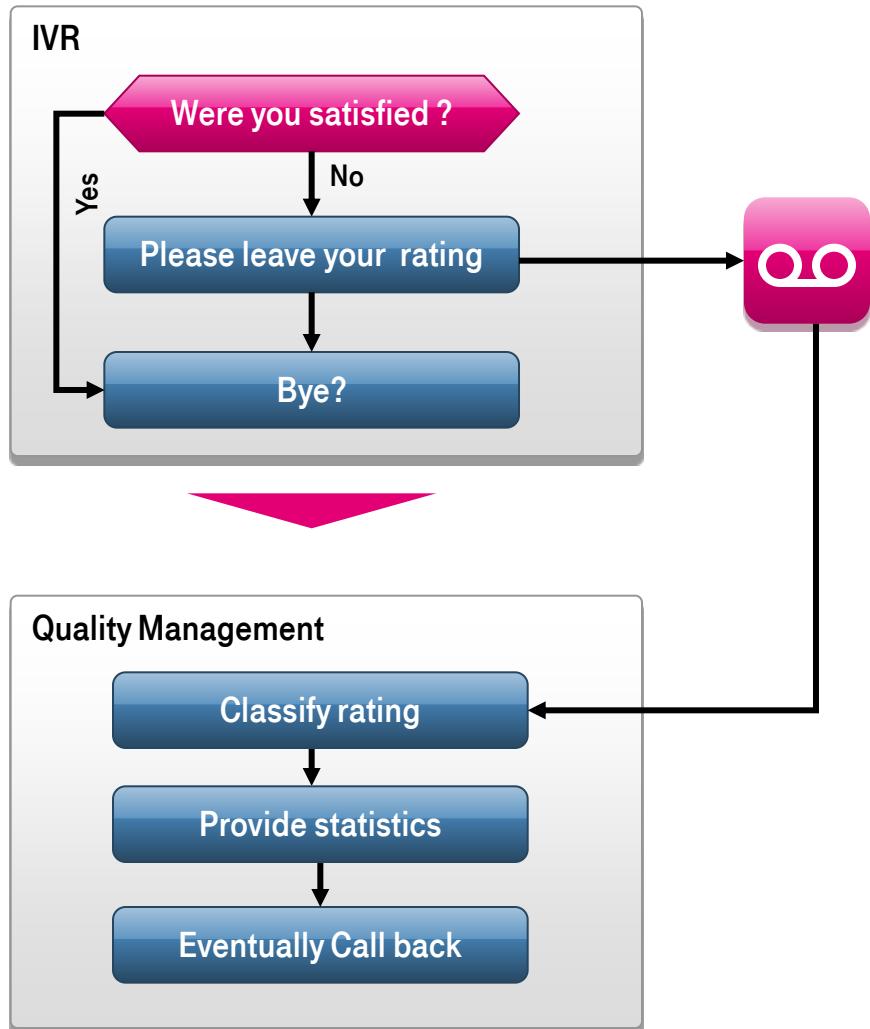
Scenario quality management. Manual evaluation of customer interviews...

Customer

- Is asked to leave a rating for the recent service.
- May have his rating being recorded as a voice file.

Agent

- Extracts voice file from text.
- Classifies customer utterance and forwards result to statistics.
- Eventually calls back in urgent cases.



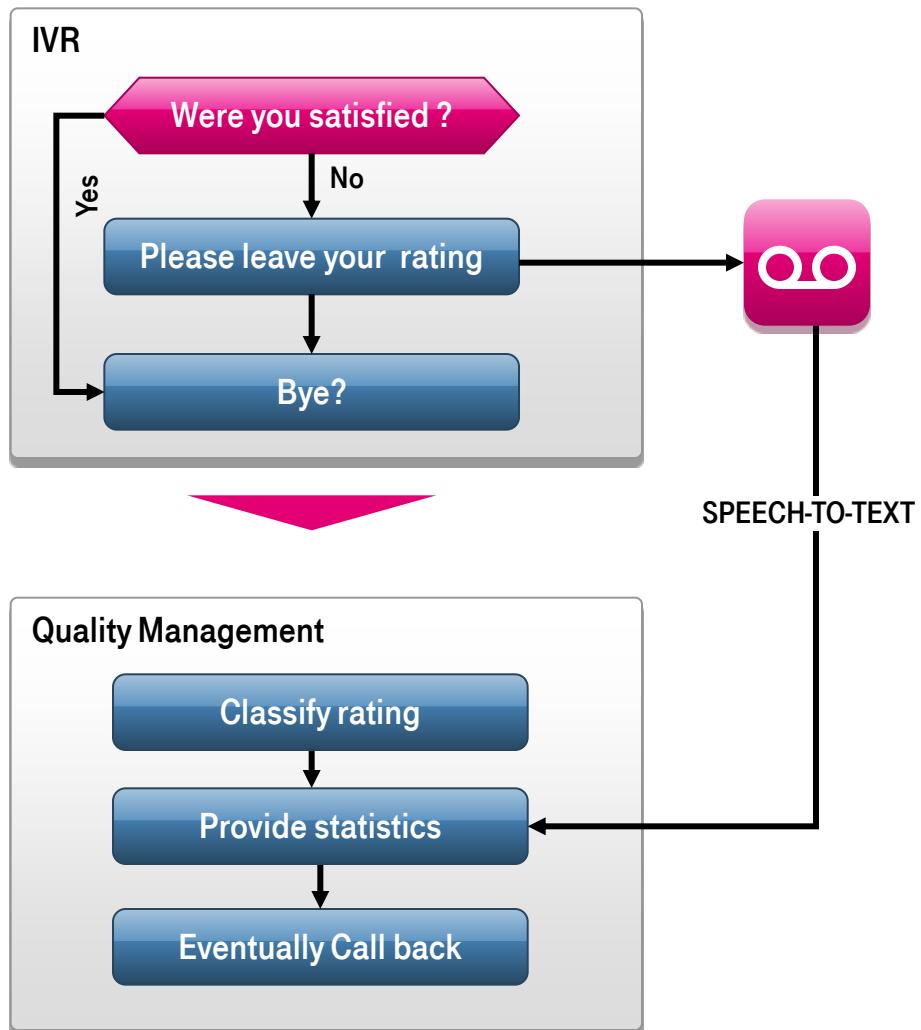
Scenario quality management. ... can be reduced by speech-to-text.

Customer

- Is asked to leave a rating for the recent service.
- May have his rating being recorded as a voice file.

Agent

- Checks transcribed text and classifications and eventually corrects them.
- Forwards result to statistics.
- Eventually calls back in urgent cases.



3. Technology.

Technology.

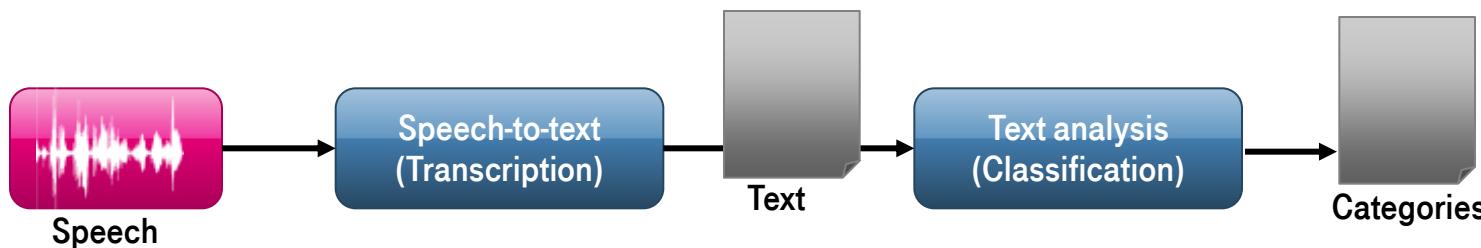
Combine speech-to-text with text classification.

Speech-to-semantics

- Speech-to-text engine with optimized SLM.
- Text Analysis (various options) and classification.

Applications

- Voice box / call-back
- Quality management



Speech-to-text

- Is based on statistical language models (SLM).
- Requires huge amount of transcribed training data for the given domain.
- Advanced tuning may be semi-automatic.
- SLMs have advantages compared to rule-based grammars for longer utterances.

Text analysis

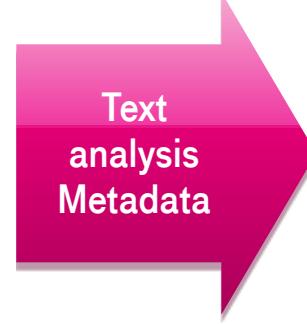
- Text classification software, assigns text to categories.
- Combines statistical classifiers (based on open source WEKA toolkit) with classification based on keyword counting.
- Keyword counting helps for stability and within sparse data situations.



Technology. How the agent should encounter it?

Example for an automatically transcribed voice message (in German).

„... Es sind mehrere Sachen ich hab ihnen vor zwei bis drei Wochen circa einen Brief geschickt und noch keine Antwort ähm also ich hatte eine Tarif also da ich nicht wusste das es für diese drei Monate auch was kostet ähm bitte stornieren danke auch ...“



Keywords, classification, and further data help the agent.

Call-back #: 06151-937999

Category: Bill

Text

Es sind mehrere Sachen ich hab ihnen vor zwei bis drei Wochen circa einen Brief geschickt und noch keine Antwort ähm also ich hatte eine **Tarif** also da ich nicht wusste das es für diese drei Monate auch **was kostet** ähm also bitte **stornieren** danke auch.

Speech-to-text alone is not sufficient!



Technology. Enhancing language models while saving Deutsche Telekom data.

In-house development and tuning of STT

- Models are optimized with live data.
- Live audio data remain inside Deutsche Telekom computing centers.

Deutsche Telekom data for Telekom services

- Data privacy requirements are respected.
- Tuned language models remain a Deutsche Telekom USP.



*WRR = word recognition rate

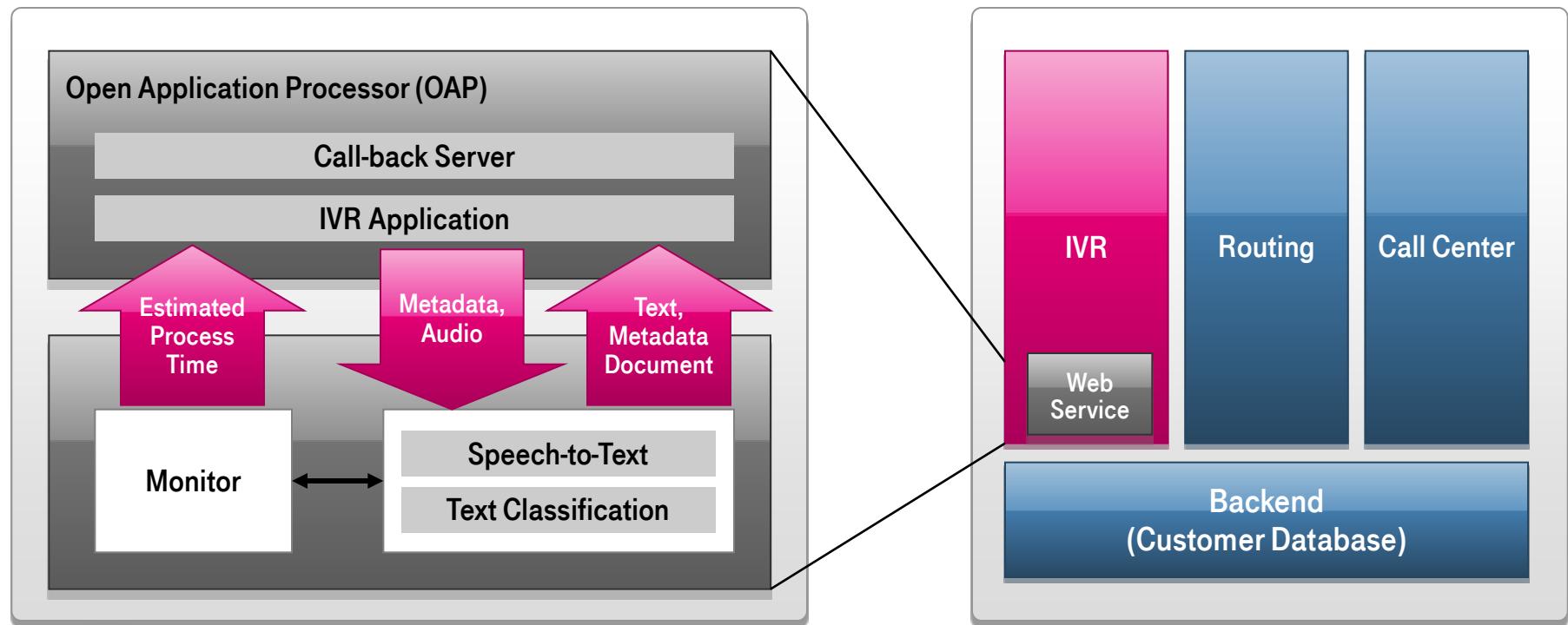


Technology.

How to integrate speech-to-text?

Integration into existing voice platform

- Speech-to-text (STT) and classification available as web service.
- A callback server inside the IVR controls STT.



Technology.

Customer is not aware of step-by-step automation.

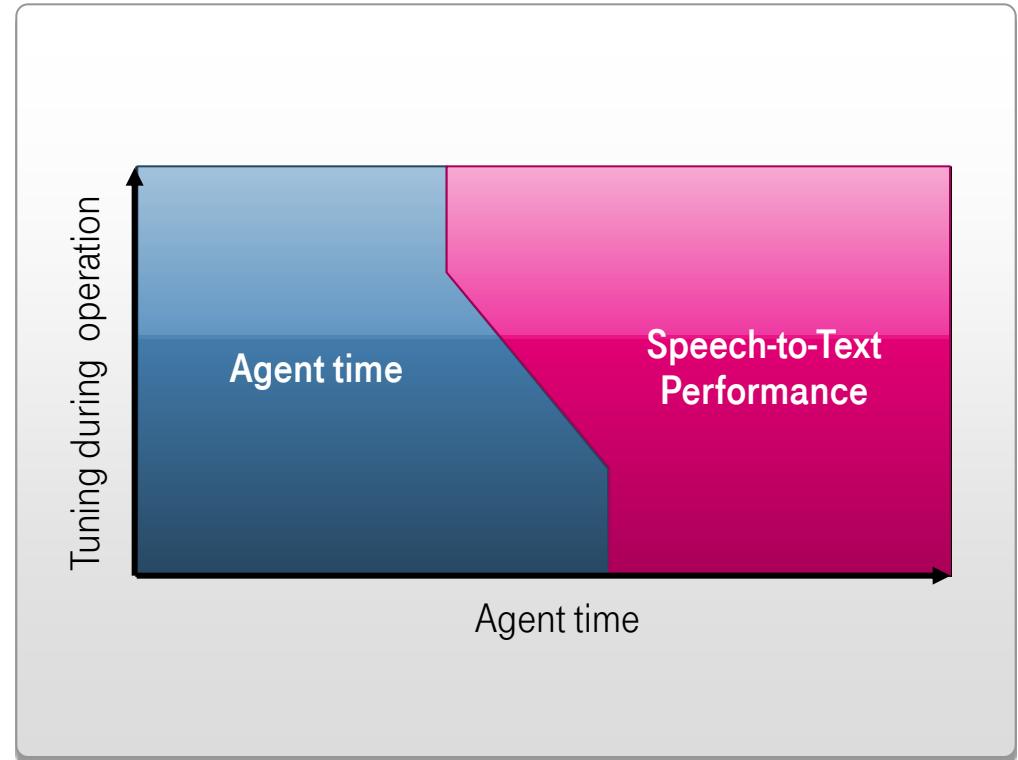
Tuning during operation

Worst case

- STT performs sub-optimally.
- Agent has little information.
- Agent needs more time.
- Customer is served anyway.

Tuning with real-life data

- Agents save time.
- Customer is served as before.



Technology.

How to make tuning of SLM cheaper?

Challenges

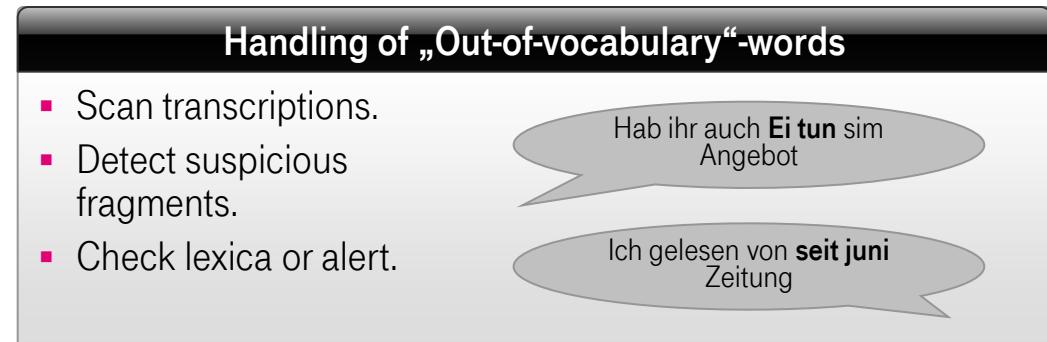
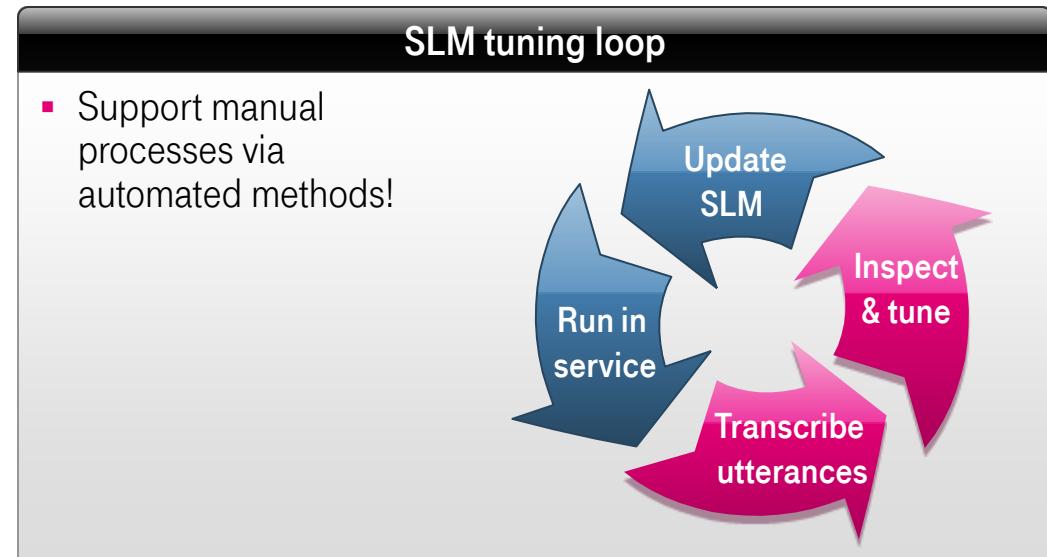
- Iterated data collection and tuning
 - Manual process, expensive.
- Out-of-vocabulary words
 - How to detect and handle them?

Semi-automatic methods

Co-operation with CMU*:

- Self supervision: guarantee performance in the absence of manual transcriptions.
- Inspection: automatically select areas for manual transcription and checking.
- Maintenance: Automatically update the language models using ASR and Web data.

* CMU: Carnegie Mellon University, Pittsburgh, USA.

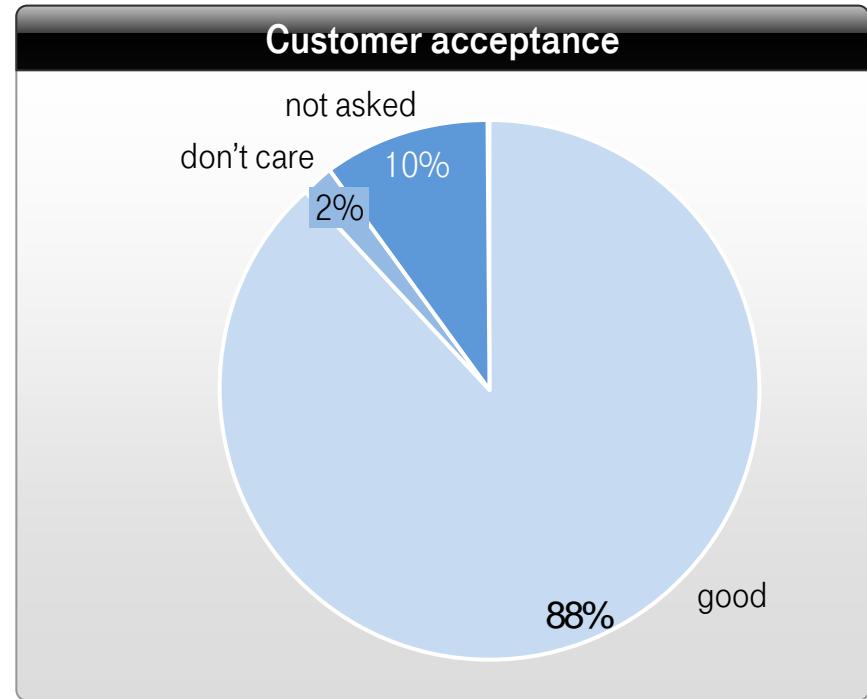
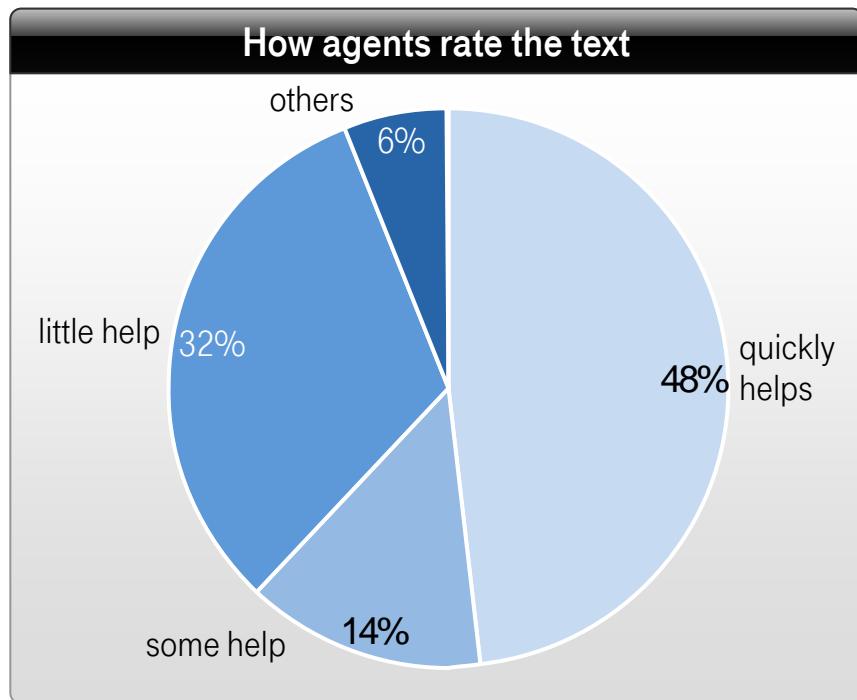


4. Results and conclusion.

Scenario call-back. Pilot customers are happy.

Summary of results after pilot with initial phase of tuning.

- Text helps agents, yet needs tuning, categorization would help further.
- Customers are happy being called back.



Scenario quality management. Speech technology proves its expected benefit.

Two ways of measuring performance

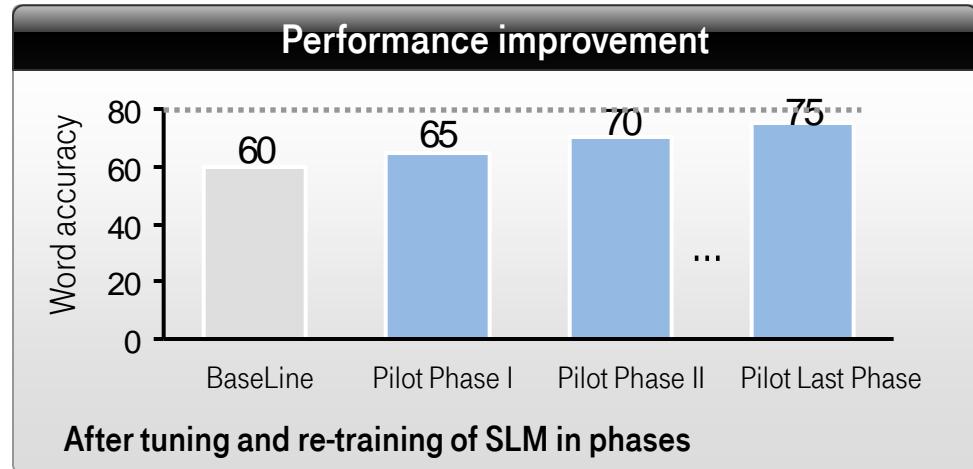
- Text quality was rated by inspector.
- Word accuracy was measured.

Results

- Text quality was rated high by inspector.
- Bad text often due to bad audio quality (e.g. background speech).
- Word accuracy of ca. 75% matches previous test results.

Inspector's suggestions for improvement

- Try to separate foreground and background speech.
- Display recognition confidences in text.
- Display punctuation or pauses in speech.
- Categorization would help.



Conclusion.

State-of-the-art

- STT is ready for applications, yet will not deliver 100%-solutions in the near future.
- Without further use of semantics, recognition in open domains will most likely not exceed 90%.
- Post-processing of the text opens options for improvement.

Solutions

- STT module is available as a web service. Thus, it suites to various applications and is easy to integrate.
- The software has proved to be stable.
- A training environment allows for administration & maintenance of speech data and language models.

Next steps

- More training with more data.
- Display confidences and punctuation in transcribed text.
- Try to detect out-of-vocabulary words.



Thank you.

• T

Life is for sharing.