



# The Road to Voicemail-to-Text in Germany

Dr. Joachim Stegmann  
Telekom Innovation Laboratories  
Mobile Voice Conference, March 14, 2012

Life is for sharing.



# Contents.

## Introduction

Legal and privacy impacts

Optimization of German language model

Results of alpha trial

Conclusion and next steps

# Introduction.

## Voice-to-text is an enabler for innovative service offerings.

### Key facts

#### Voice-to-text (VTT)

- Automatic transcription of voice files into text is enabler for innovative service offerings like Voicemail-to-text, message dictation.



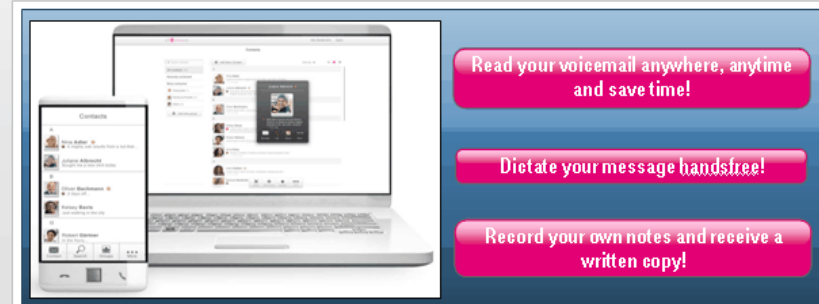
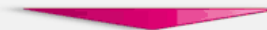
#### Read your voicemail

- Anywhere, anytime - in a noisy environment or in a meeting!
- Spontaneous, telephone bandwidth (3.4 kHz) speech.

#### Dictate your text message hands-free

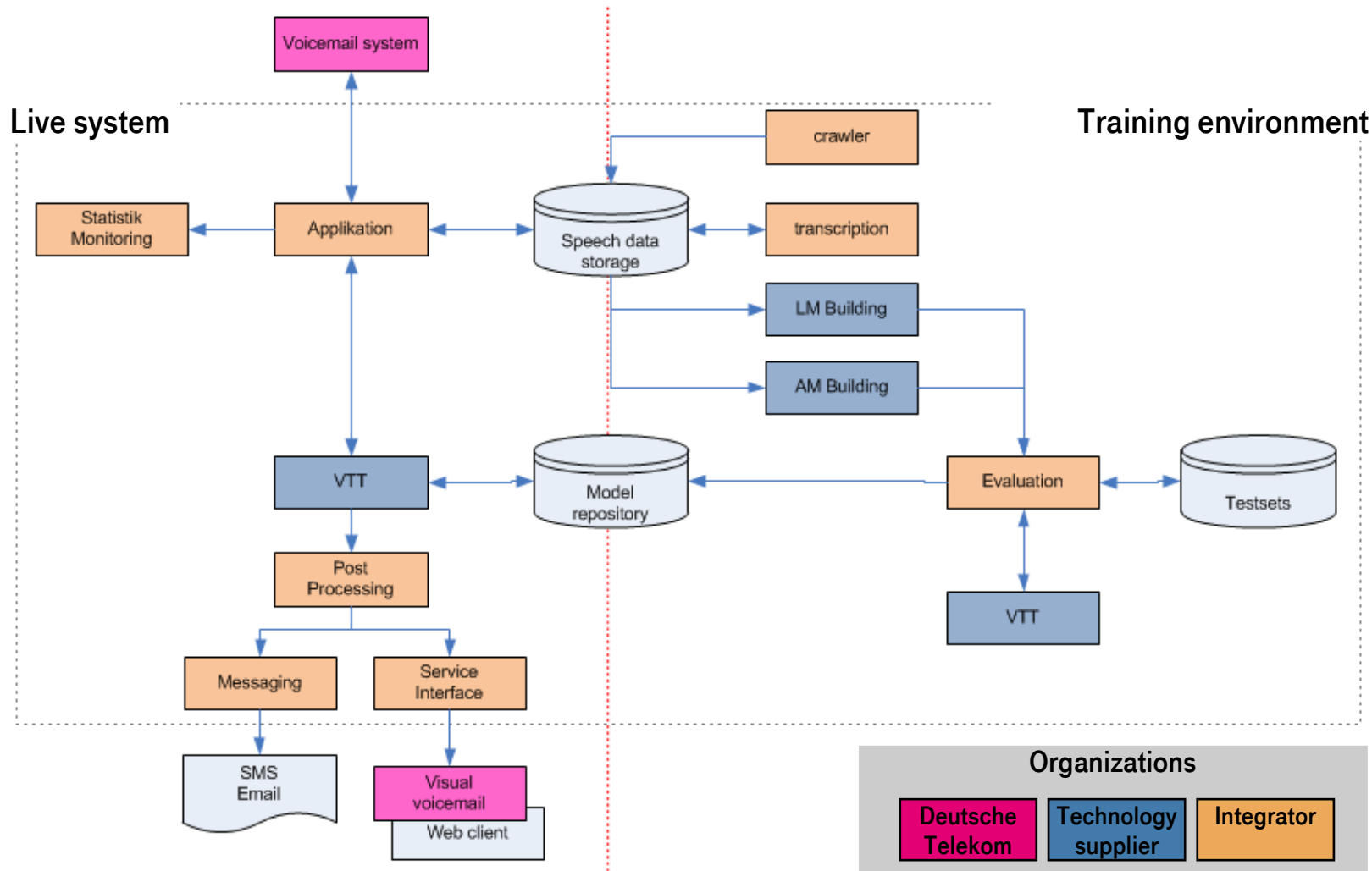
- On the move, while driving your car.
- Mostly structured and wideband (7 kHz) speech.

### Current scenario



# Introduction.

Processes and roles are defined in the service architecture.



# Contents.

Introduction

Legal and privacy impacts

Optimization of German language model

Results of alpha trial

Conclusion and next steps

# Legal and privacy impacts.

German law and specific DTAG requirements must be considered.

| Constraints   | Training with man.transcription  | Operation w/o man.transcription |
|---|--|---------------------------------|
| Special prompt for A-party concerning "Voice-to-Text" functionality                                 | ●<br>(with indication of man. transcription)   | ●                               |
| Active consent of A-party with choice to decline (§88 TKG*)   | ●  | —                               |
| Consent of B-party accordant to §107 TKG* e.g. in the context of AGB** or a data privacy statement. | ●  | ●                               |
| Transcription by third party (§88 TKG)  | ○<br>With following constraints:<br><ul style="list-style-type: none"> <li>▪ Special prompt for A-party concerning transcription by third party</li> <li>▪ Anonymization/encryption</li> <li>▪ Org. controls (audits...)</li> <li>▪ Only in Germany</li> </ul> | —                               |
| Data storage  | Only in Germany  | Only in Germany                 |
| Operation   | Only by staff of DTAG  | Only by staff of DTAG           |

Legend: ● = required, ○ = possible with constraints, — = not applicable

**On-premise solution meets best the privacy and security requirements of Deutsche Telekom. Aspects like exclusivity of DTAG speech corpora have to be considered, too.**



# Contents.

Introduction

Legal and privacy impacts

Optimization of German language model

Results of alpha trial

Conclusion and next steps.

# Optimization of German language model. Status at project start after first data collection.

## Status after project start with first data collection

### Data collection and optimization

- 50,000 'artificial' voice mails of 5,000 speakers are collected with an agency and transcribed for optimization of the basic German LMs of two technology suppliers.
- 1,000 separate voice mails have been used for evaluation.



### Results from evaluation February 2011

|           | Supplier | WER   |
|-----------|----------|-------|
| Basic LMs | 1        | 24,1% |
|           | 2        | 32,4% |
|           | 3        | -     |

|             | Supplier | WER   |
|-------------|----------|-------|
| Adapted LMs | 1        | 23,1% |
|             | 2        | 25,7% |
|             | 3        | -     |

An initial system can be derived from a general LM with around 50,000 transcribed voicemails





# Optimization of German language model. Current status after alpha trial.

## Status after alpha trial and evaluation with real data

### Data collection and optimization

- 1,000 real voice mails (116 speakers) were stored and transcribed during the alpha trial for estimating the real performance of the technology suppliers.

### Results from evaluation February 2012

|                                     | Supplier | WER   |
|-------------------------------------|----------|-------|
| Evaluation of all 1,000 voice mails | 1        | 29.9% |
|                                     | 2        | 37.6% |
|                                     | 3        | 38.3% |

|   | Supplier | WER   |
|---|----------|-------|
| Evaluation of reduced set of 611 voice mails* | 1        | 22.6% |
|   | 2        | 28.9% |
|   | 3        | 18.8% |

- \*Some suppliers refuse to process voice mails with a low confidence level. The second test includes only those voice mails that have been processed by all suppliers.

**Comparison of different vendors heavily depends on the test set, WER is not a reliable measure.**



# Optimization of German language model. Include post processing and new features.

## Next steps

### Speech recognition

*Significant reduction of WER expected for the next weeks!*

- Lexicon enlargement & new acoustic model(s)

### Post-processing and presentation

*Improve understandability and readability!*

- Optimize presentation of numbers, date, time, URLs and email addresses.
- Add punctuation (commas, full stops, exclamation marks, question marks), highlighting and line breaks.

### New Features

*Improve usability!*

- Tag words with low recognition confidence and show word alternatives (n-best lists).
- Start audio output of your voicemail at any position of the text.



# Contents.

Introduction

Legal and privacy impacts

Optimization of German language model

Results of alpha trial

Conclusion and next steps

# Alpha trial.

Customers have used the voicemail app in daily life.

## Test setup, test support and test analysis

### User tests

- Field work: Q4, 2011
- Laboratory tests: 166 test users have to solve tasks in a laboratory environment at four locations all over Germany to eliminate the influence of dialects.
- Home tests: afterwards test users install the apps on their own smart phones and use them in daily life.
- Age (18-34 / 35-64 years, 50% each), gender (50% each).

### Test setup & support

- Questionnaire & test scenarios specified.
- Prototypes & handouts for mailbox app.

### Test analysis

- Expert transcriptions of all audio data collected during the tests.
- Calculation of test user specific WERs.



# Alpha trial.

The subjective results show high customer satisfaction.

## Results Voicemail App



- After 3 weeks 'need fulfilment' has significantly increased up to 52% (!) for top 2 boxes (rating 4 or 5) of a scale from 1 (does not fulfil) to 5 (does fulfil).
- At 80% for top-2-boxes 5 and 4 for a 1 (not unique at all) to 5 (very unique) scale, 'uniqueness' is very high!
- Also high scores for 'likability' (55% for top-2-boxes).
- 42% are very or extremely interested even with the current VTT accuracy!
- Top-2-box purchase intention (1 for "would definitely not buy it" and 5 for "would definitely buy it") has significantly increased to 35%!

**The sum of benefits of the App is strong enough to overcome the accuracy issue. Results are good enough to consider a beta launch with the current engine(s) accuracy.**



# Contents.

Introduction

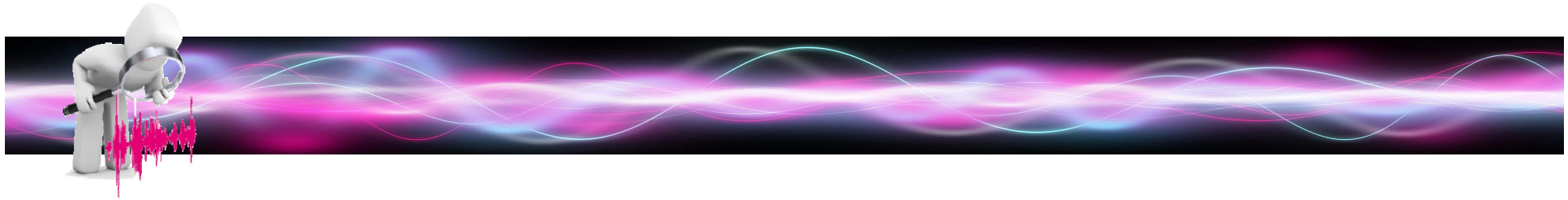
Legal and privacy impacts

Optimization of German language model

Results of alpha trial

Conclusion and next steps

## Conclusion and next steps.



- The alpha trial demonstrates the market maturity of voice-to-text (VTT) technology for the use case voicemail-text although the level of accuracy related to WER is not very high.
- Further improvement of transcription accuracy has to be in focus. For an improvement of readability, understandability and usability the post-processing is also essential.
- On-premise solution meets best the privacy and security requirements of DTAG. Additionally, aspects like exclusivity of DTAG speech corpora have to be considered.
- Thus DTAG is recommended to act as the collector of speech data. Training of German language models might also be covered by DTAG in a mid term.
- Next steps: Planning of an open beta trial in 2012.



Thank you for your attention.

