



ulm university

universität
uulm

Speech-Emotion Recognition in Adaptive Dialogue Systems

Promotionskolloquium

Johannes Pittermann | Juli 2008 | Institut für Informationstechnik

Motivation

- ▶ Emotional intelligente Dialogsysteme

- ▶ Zusätzliche Komponenten

- ▶ Robuste Erkennung von Emotionen

- ▶ Konsistentes Modell zur Anpassung des Dialogs

Motivation

- ▶ Emotional intelligente Dialogsysteme



- ▶ Zusätzliche Komponenten
 - ▶ Robuste Erkennung von Emotionen
 - ▶ Konsistentes Modell zur Anpassung des Dialogs

Motivation

- ▶ Emotional intelligente Dialogsysteme



- ▶ Zusätzliche Komponenten
 - ▶ Robuste Erkennung von Emotionen
 - ▶ Konsistentes Modell zur Anpassung des Dialogs

Motivation

- ▶ Emotional intelligente Dialogsysteme



- ▶ Zusätzliche Komponenten

- ▶ Robuste Erkennung von Emotionen
- ▶ Konsistentes Modell zur Anpassung des Dialogs

Motivation

- ▶ Emotional intelligente Dialogsysteme



- ▶ Zusätzliche Komponenten
 - ▶ Robuste Erkennung von Emotionen
 - ▶ Konsistentes Modell zur Anpassung des Dialogs

Motivation

- ▶ Emotional intelligente Dialogsysteme



- ▶ Zusätzliche Komponenten
 - ▶ Robuste Erkennung von Emotionen
 - ▶ Konsistentes Modell zur Anpassung des Dialogs

Inhalt

- ▶ Einführung
- ▶ Sprach-Emotionserkennung
- ▶ Dialogmodellierung
- ▶ Zusammenfassung & Ausblick

Inhalt

- ▶ Einführung
- ▶ Sprach-Emotionserkennung
- ▶ Dialogmodellierung
- ▶ Zusammenfassung & Ausblick

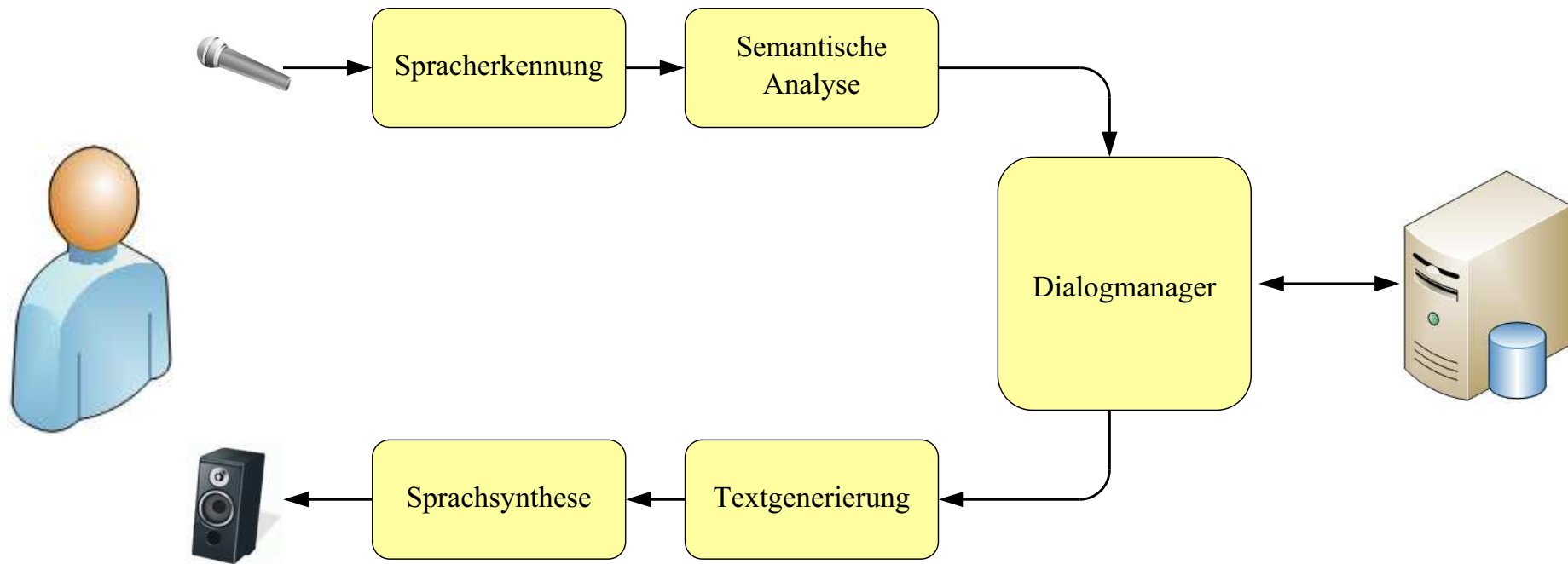
Inhalt

- ▶ Einführung
- ▶ Sprach-Emotionserkennung
- ▶ Dialogmodellierung
- ▶ Zusammenfassung & Ausblick

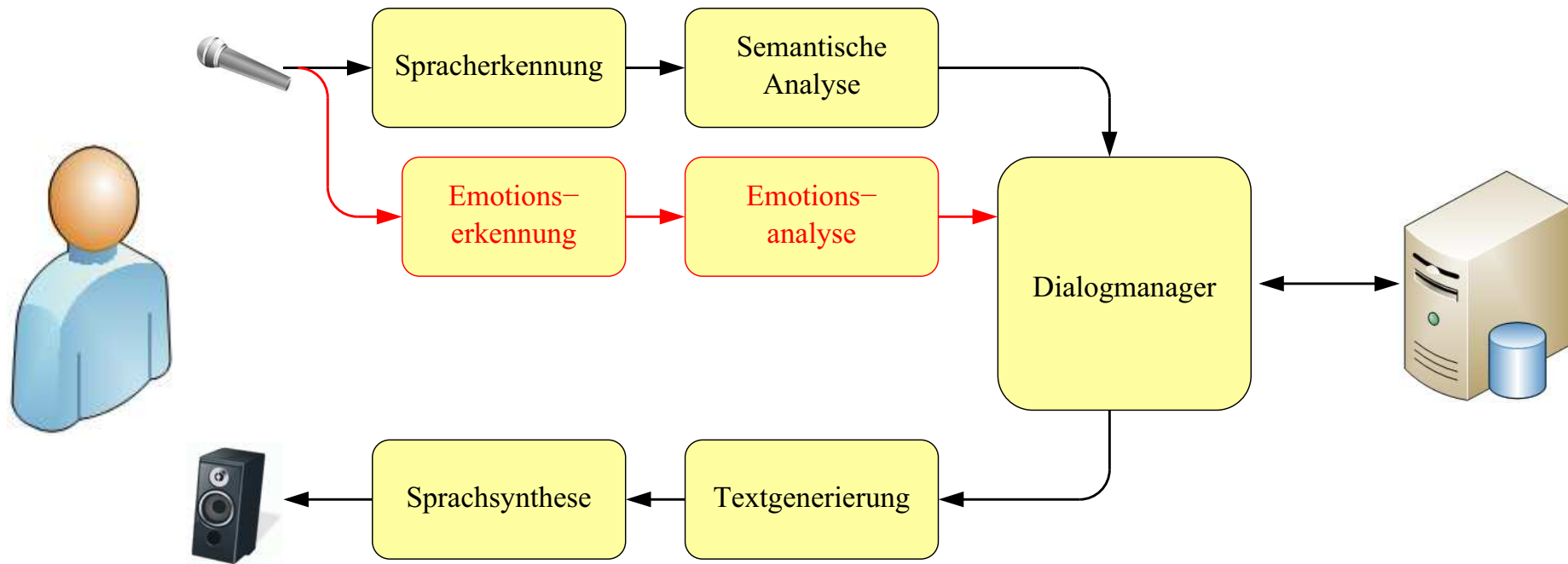
Inhalt

- ▶ Einführung
- ▶ Sprach-Emotionserkennung
- ▶ Dialogmodellierung
- ▶ Zusammenfassung & Ausblick

Sprachdialogsysteme



Adaptive Sprachdialogsysteme



Berücksichtigte Emotionen

- ▶ Keine einheitliche Kategorisierung \Rightarrow *Primäremotionen*
- ▶ Emotionale Sprachdatenbank (TU Berlin)
 - ▶ 7 Emotionen:
Angst, Ekel, Freude, Langeweile, Neutral, Trauer, Wut
 - ▶ 10 Sprecher
 - ▶ 10 Sätze

Trauer

Wut

Berücksichtigte Emotionen

- ▶ Keine einheitliche Kategorisierung \Rightarrow *Primäremotionen*
- ▶ Emotionale Sprachdatenbank (TU Berlin)
 - ▶ 7 Emotionen:
Angst, Ekel, Freude, Langeweile, Neutral, Trauer, Wut
 - ▶ 10 Sprecher
 - ▶ 10 Sätze

Trauer

Wut

Berücksichtigte Emotionen

- ▶ Keine einheitliche Kategorisierung \Rightarrow *Primäremotionen*
- ▶ Emotionale Sprachdatenbank (TU Berlin)
 - ▶ 7 Emotionen:
Angst, Ekel, Freude, Langeweile, Neutral, Trauer, Wut
 - ▶ 10 Sprecher
 - ▶ 10 Sätze

Trauer

Wut

Berücksichtigte Emotionen

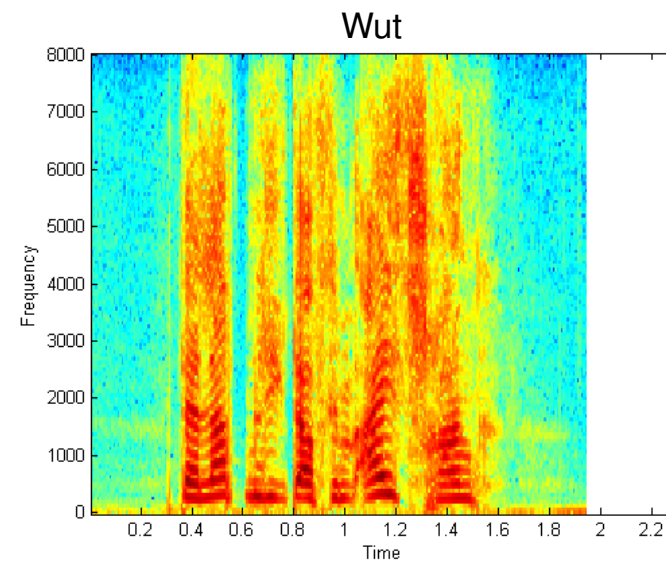
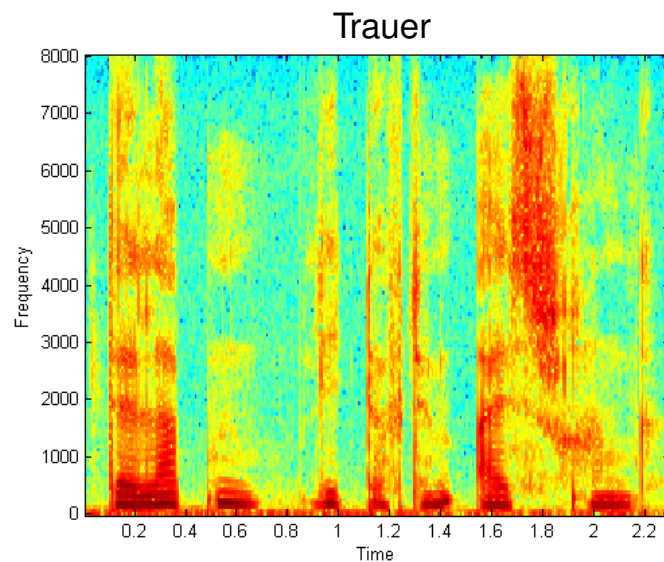
- ▶ Keine einheitliche Kategorisierung \Rightarrow *Primäremotionen*
- ▶ Emotionale Sprachdatenbank (TU Berlin)
 - ▶ 7 Emotionen:
Angst, Ekel, Freude, Langeweile, Neutral, Trauer, Wut
 - ▶ 10 Sprecher
 - ▶ 10 Sätze

Trauer

Wut

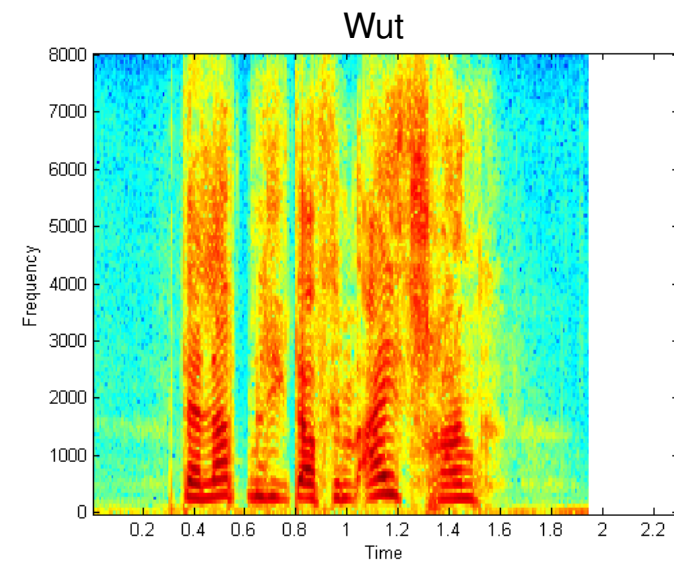
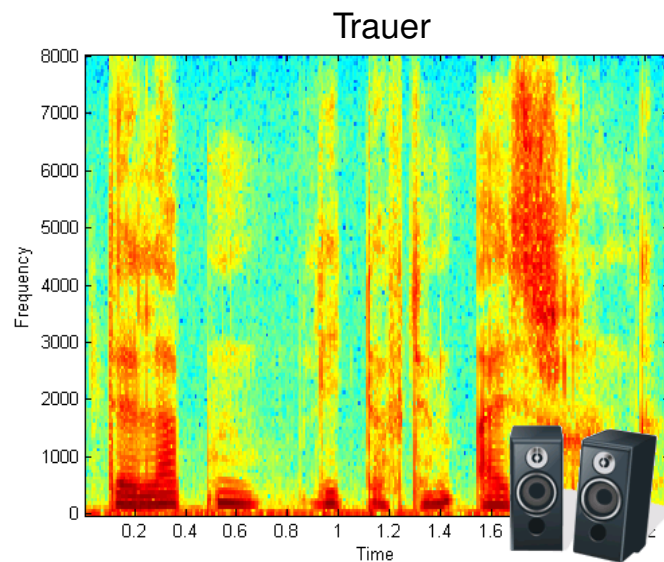
Berücksichtigte Emotionen

- ▶ Keine einheitliche Kategorisierung \Rightarrow *Primäremotionen*
- ▶ Emotionale Sprachdatenbank (TU Berlin)
 - ▶ 7 Emotionen:
Angst, Ekel, Freude, Langeweile, Neutral, Trauer, Wut
 - ▶ 10 Sprecher
 - ▶ 10 Sätze



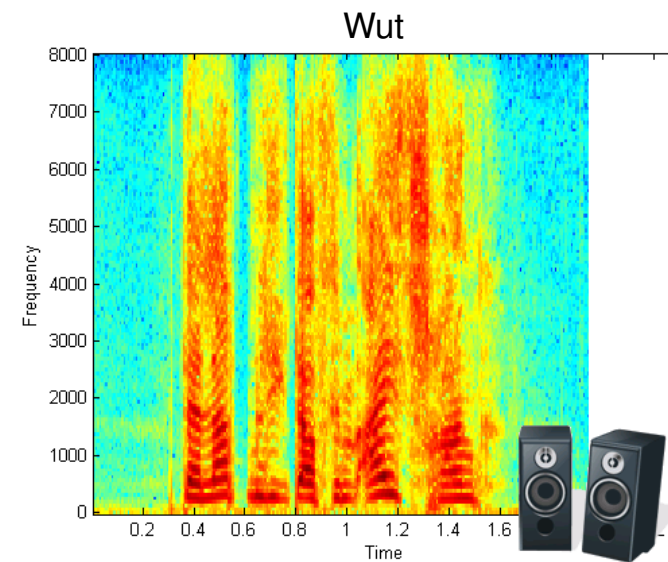
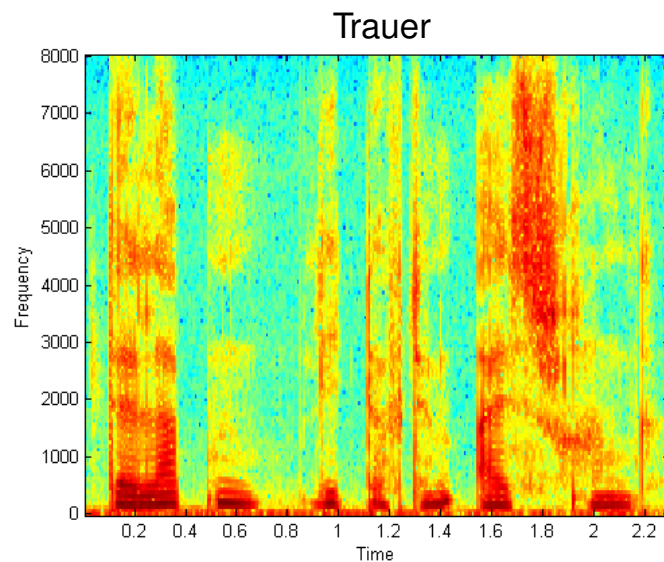
Berücksichtigte Emotionen

- ▶ Keine einheitliche Kategorisierung \Rightarrow *Primäremotionen*
- ▶ Emotionale Sprachdatenbank (TU Berlin)
 - ▶ 7 Emotionen:
Angst, Ekel, Freude, Langeweile, Neutral, Trauer, Wut
 - ▶ 10 Sprecher
 - ▶ 10 Sätze

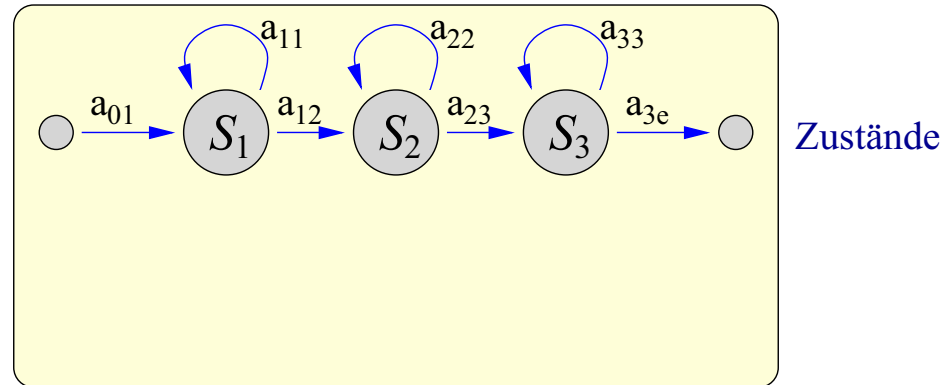


Berücksichtigte Emotionen

- ▶ Keine einheitliche Kategorisierung \Rightarrow *Primäremotionen*
- ▶ Emotionale Sprachdatenbank (TU Berlin)
 - ▶ 7 Emotionen:
Angst, Ekel, Freude, Langeweile, Neutral, Trauer, Wut
 - ▶ 10 Sprecher
 - ▶ 10 Sätze

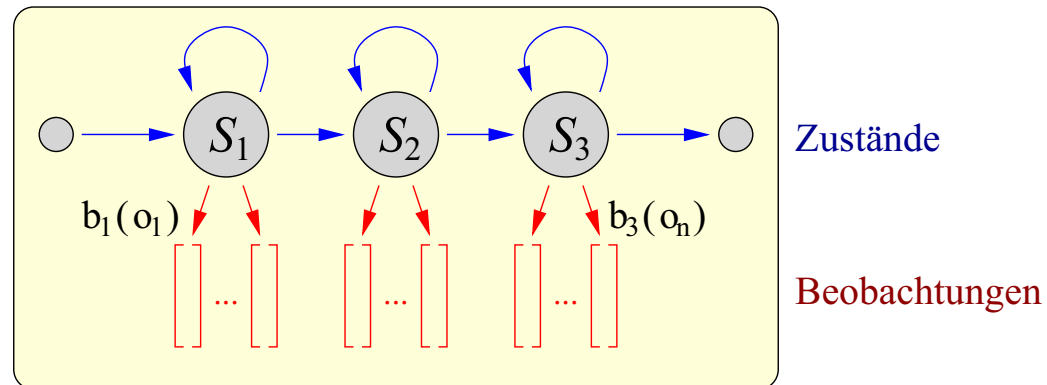


Standard-HMM Emotionserkennung



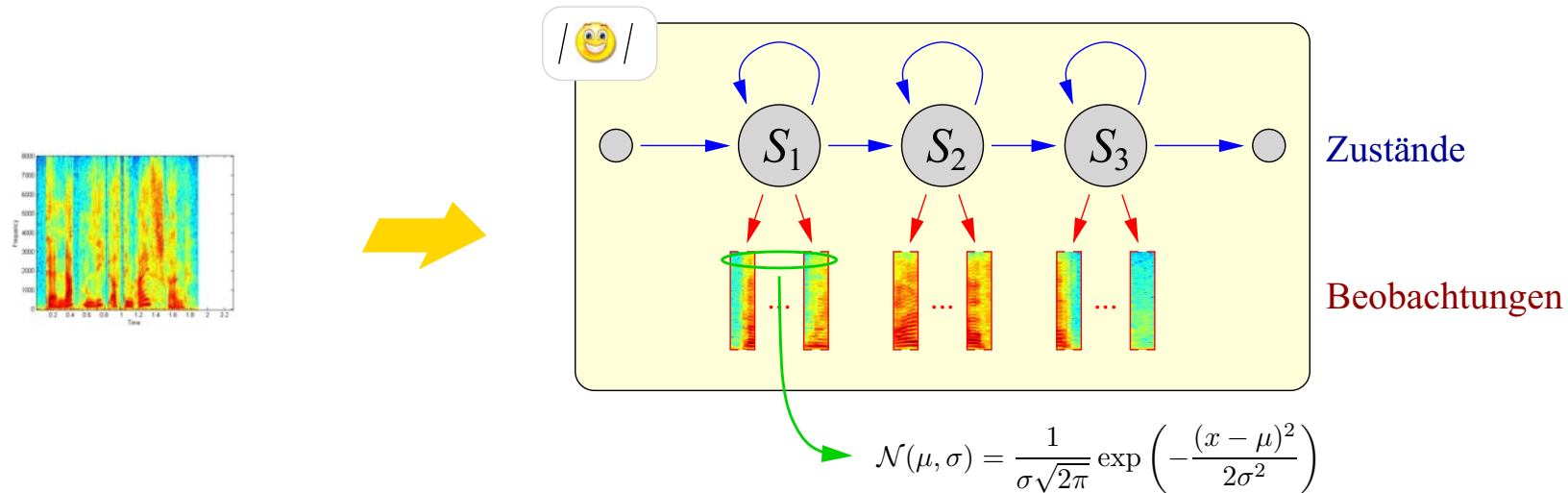
- ▶ Hidden Markov Models zur akustischen Modellierung
- ▶ Akustische Merkmale: Tonhöhe, Formanten, Intensität, Qualitätsmerkmale
- ▶ Unterschiedliche Modelle für weibliche und männliche Sprecher

Standard-HMM Emotionserkennung



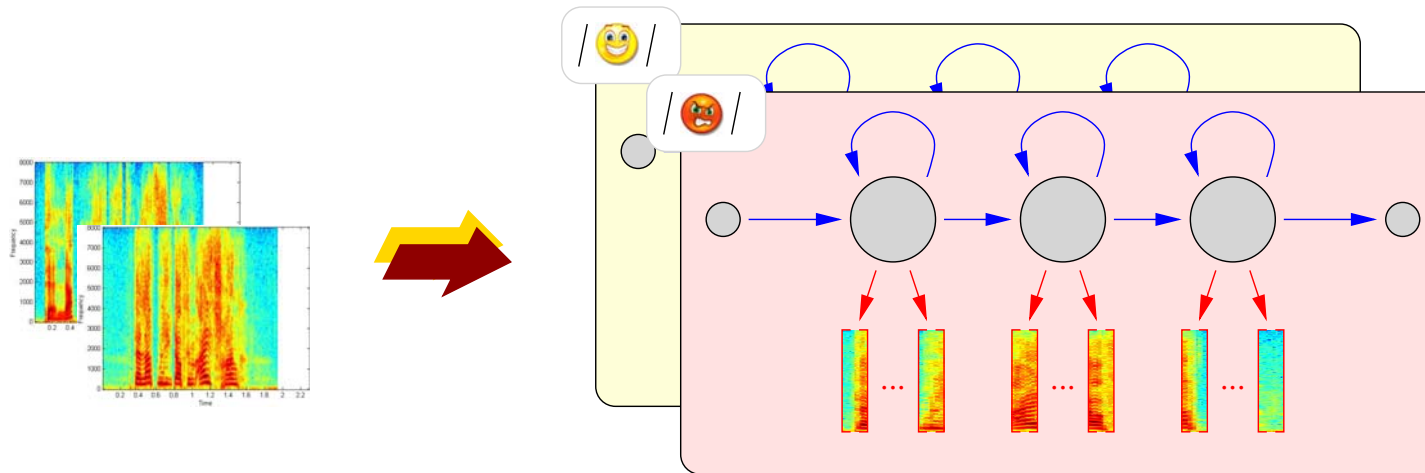
- ▶ Hidden Markov Models zur akustischen Modellierung
- ▶ Akustische Merkmale: Tonhöhe, Formanten, Intensität, Qualitätsmerkmale
- ▶ Unterschiedliche Modelle für weibliche und männliche Sprecher

Standard-HMM Emotionserkennung



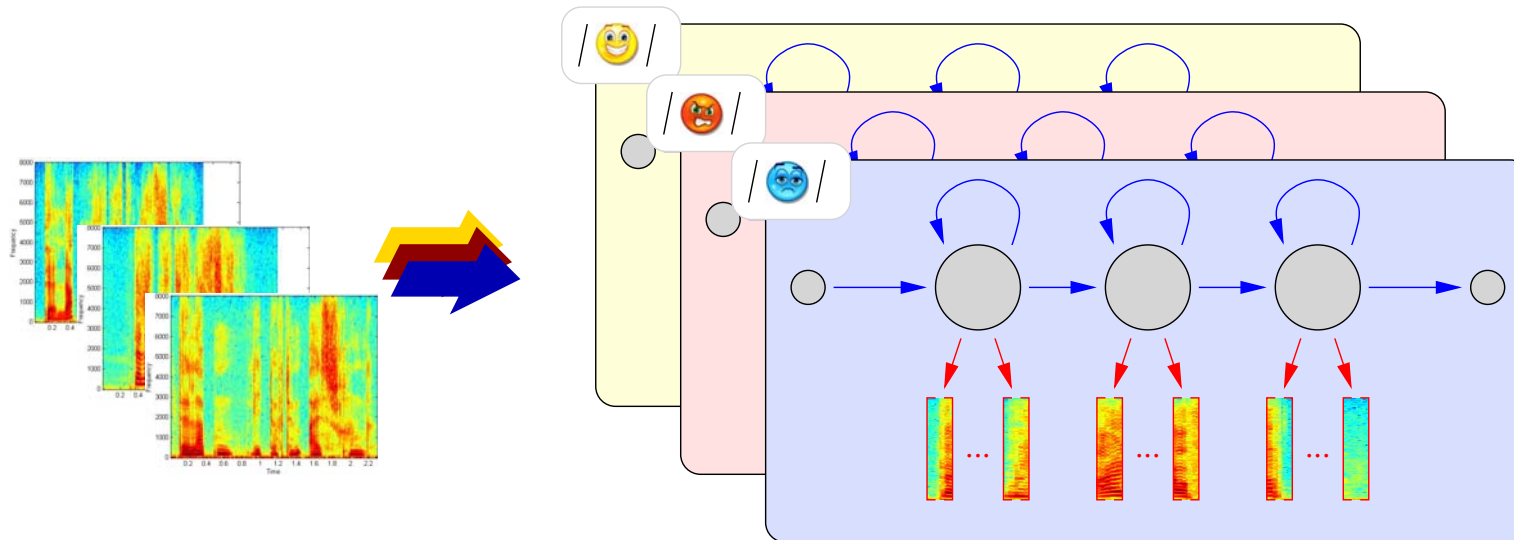
- ▶ Hidden Markov Models zur akustischen Modellierung
- ▶ Akustische Merkmale: Tonhöhe, Formanten, Intensität, Qualitätsmerkmale
- ▶ Unterschiedliche Modelle für weibliche und männliche Sprecher

Standard-HMM Emotionserkennung



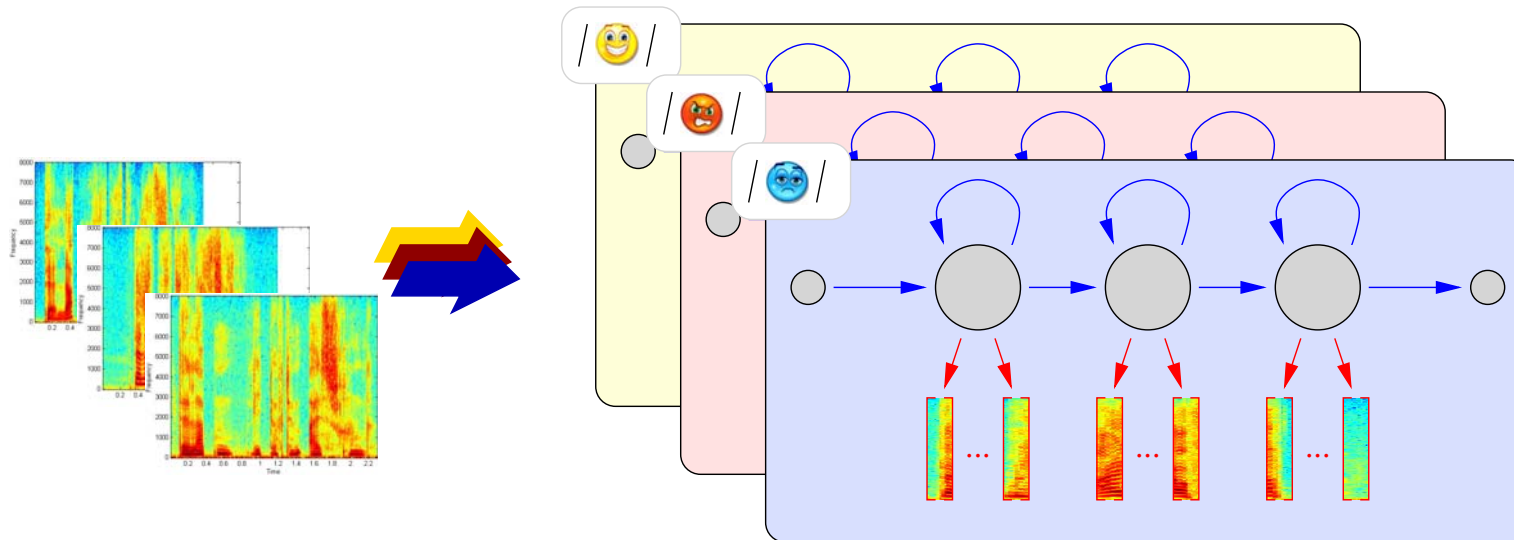
- ▶ Hidden Markov Models zur akustischen Modellierung
- ▶ Akustische Merkmale: Tonhöhe, Formanten, Intensität, Qualitätsmerkmale
- ▶ Unterschiedliche Modelle für weibliche und männliche Sprecher

Standard-HMM Emotionserkennung



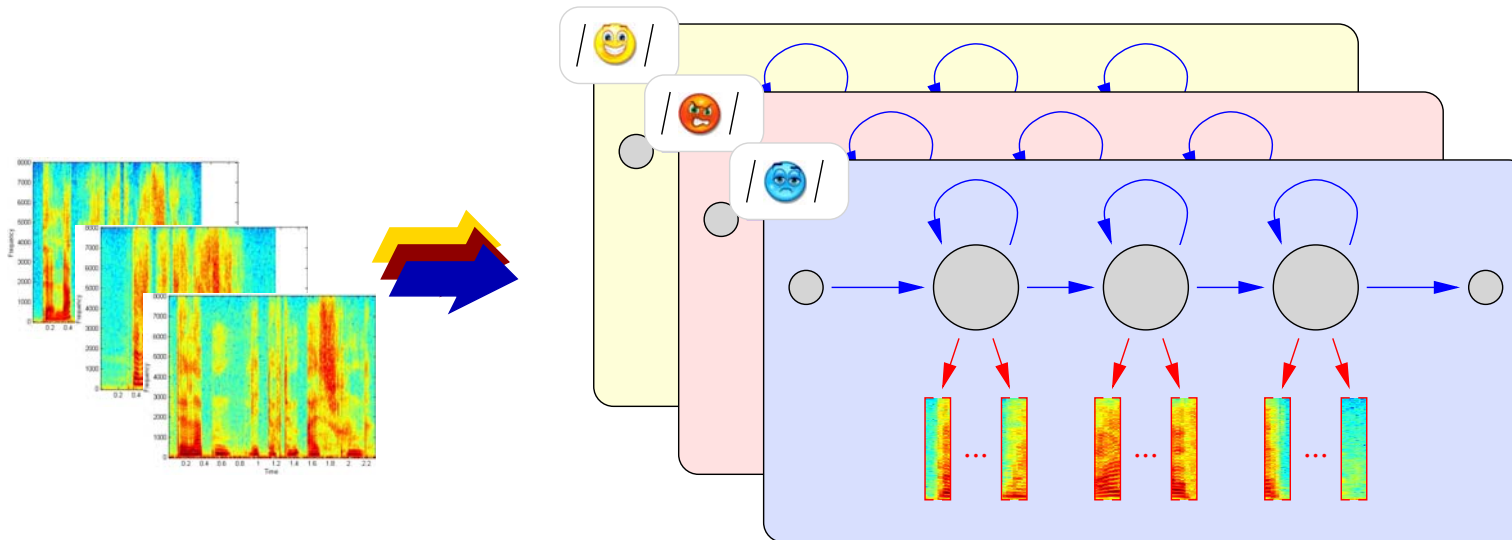
- ▶ Hidden Markov Models zur akustischen Modellierung
- ▶ Akustische Merkmale: Tonhöhe, Formanten, Intensität, Qualitätsmerkmale
- ▶ Unterschiedliche Modelle für weibliche und männliche Sprecher

Standard-HMM Emotionserkennung



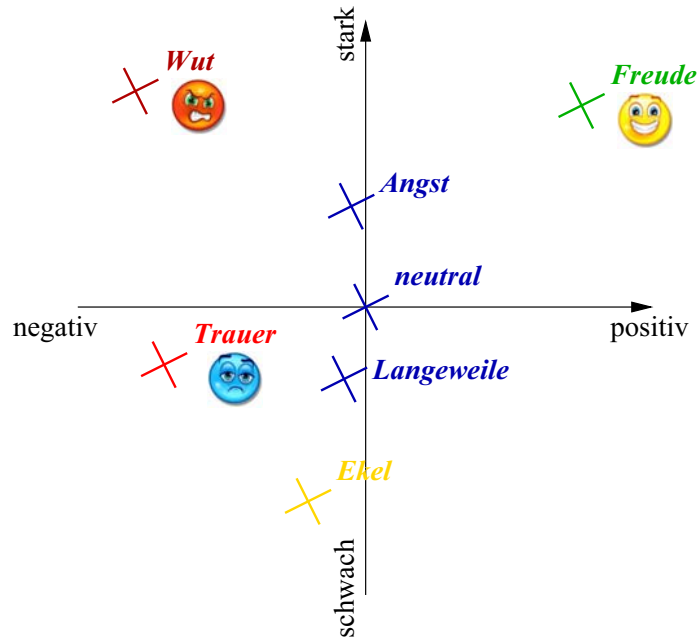
- ▶ Hidden Markov Models zur akustischen Modellierung
- ▶ Akustische Merkmale: Tonhöhe, Formanten, Intensität, Qualitätsmerkmale
- ▶ Unterschiedliche Modelle für weibliche und männliche Sprecher

Standard-HMM Emotionserkennung



- ▶ Hidden Markov Models zur akustischen Modellierung
- ▶ Akustische Merkmale: Tonhöhe, Formanten, Intensität, Qualitätsmerkmale
- ▶ Unterschiedliche Modelle für weibliche und männliche Sprecher

Emotionen als Dialogparameter



► Darstellung durch Zahlenwerte:

► Valence-Arousal-Ebene:

$$(-1, -1) \leq (v, a) \leq (1, 1)$$

► 1-dim. Emotionswert:

$$0 \leq E(U) \leq 2$$

► “Bedrohung” (threat):

$$\theta = \frac{1}{3} E(U) \cdot (D_{BS} + P_{BS} + V_I)$$

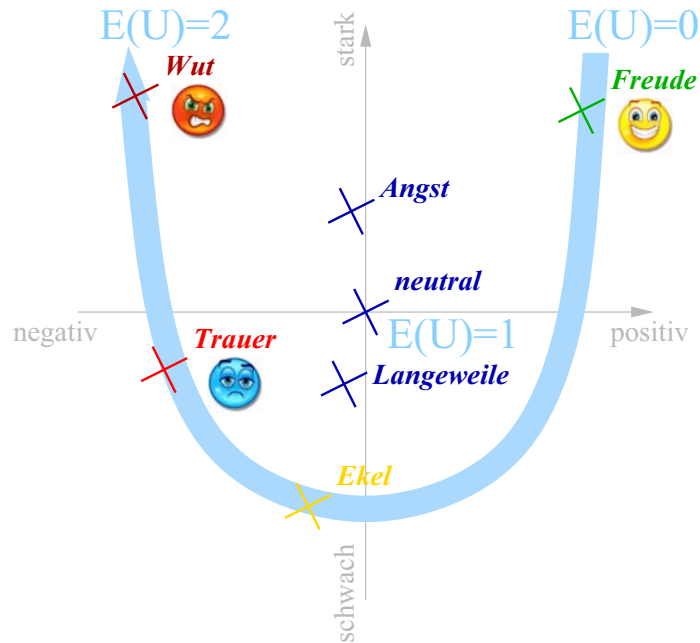
► Regelbasierte Adaption des Dialogablaufs

$E(U) = 0$: **Großartig**, *wohin soll die Reise gehen?*

$E(U) = 1$: *Wohin möchten Sie reisen?*

$E(U) = 2$: **Es tut mir Leid**, *dürfte ich bitte erfahren, wohin Sie reisen möchten?*

Emotionen als Dialogparameter



► Darstellung durch Zahlenwerte:

► Valence-Arousal-Ebene:

$$(-1, -1) \leq (v, a) \leq (1, 1)$$

► 1-dim. Emotionswert:

$$0 \leq E(U) \leq 2$$

► “Bedrohung” (threat):

$$\theta = \frac{1}{3} E(U) \cdot (D_{BS} + P_{BS} + V_I)$$

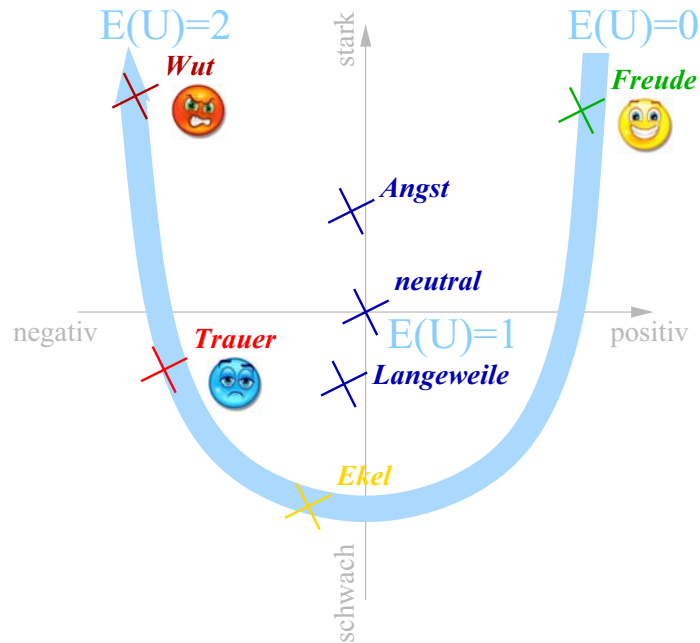
► Regelbasierte Adaption des Dialogablaufs

$E(U) = 0$: “**Großartig**, wohin soll die Reise gehen?”

$E(U) = 1$: “Wohin möchten Sie reisen?”

$E(U) = 2$: “**Es tut mir Leid**, dürfte ich bitte erfahren, wohin Sie reisen möchten?”

Emotionen als Dialogparameter



► Darstellung durch Zahlenwerte:

► Valence-Arousal-Ebene:

$$(-1, -1) \leq (v, a) \leq (1, 1)$$

► 1-dim. Emotionswert:

$$0 \leq E(U) \leq 2$$

► “Bedrohung” (threat):

$$\theta = \frac{1}{3} E(U) \cdot (D_{BS} + P_{BS} + V_I)$$

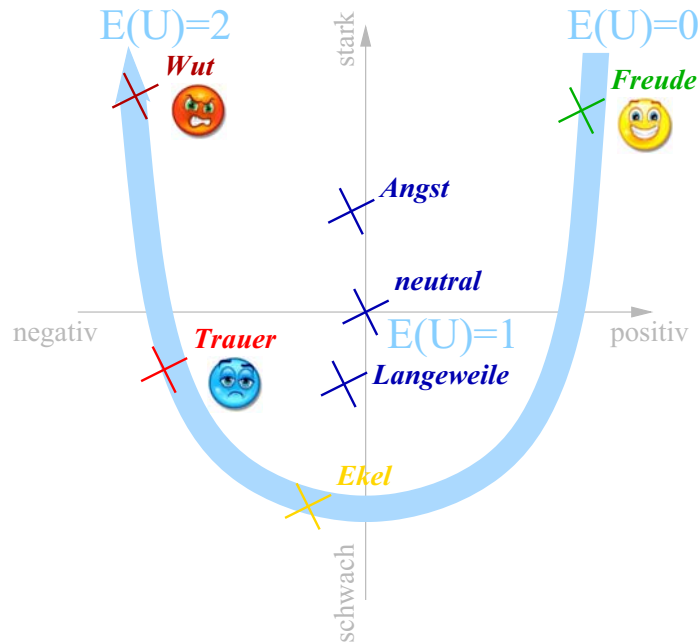
► Regelbasierte Adaption des Dialogablaufs

$E(U) = 0$: **Großartig**, *wohin soll die Reise gehen?*

$E(U) = 1$: *Wohin möchten Sie reisen?*

$E(U) = 2$: **Es tut mir Leid**, *dürfte ich bitte erfahren, wohin Sie reisen möchten?*

Emotionen als Dialogparameter



► Darstellung durch Zahlenwerte:

► Valence-Arousal-Ebene:

$$(-1, -1) \leq (v, a) \leq (1, 1)$$

► 1-dim. Emotionswert:

$$0 \leq E(U) \leq 2$$

► “Bedrohung” (threat):

$$\theta = \frac{1}{3} E(U) \cdot (D_{BS} + P_{BS} + V_I)$$

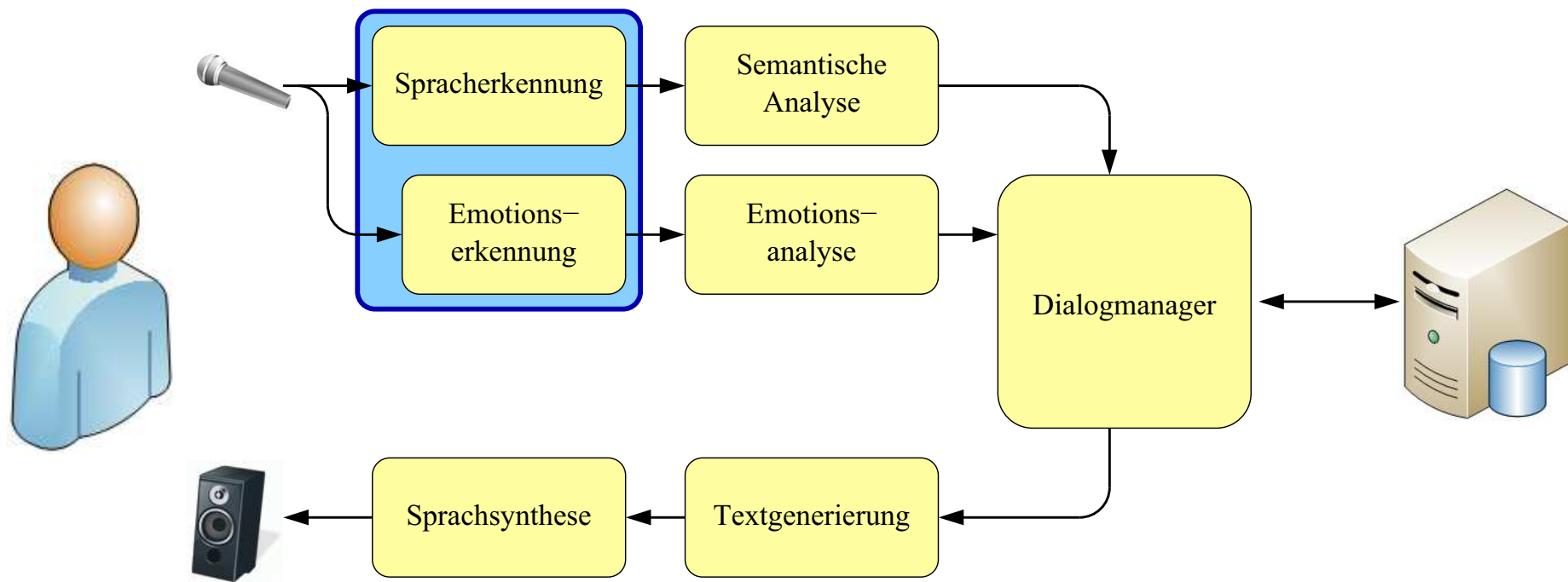
► Regelbasierte Adaption des Dialogablaufs

$E(U) = 0$: **Großartig**, *wohin soll die Reise gehen?*

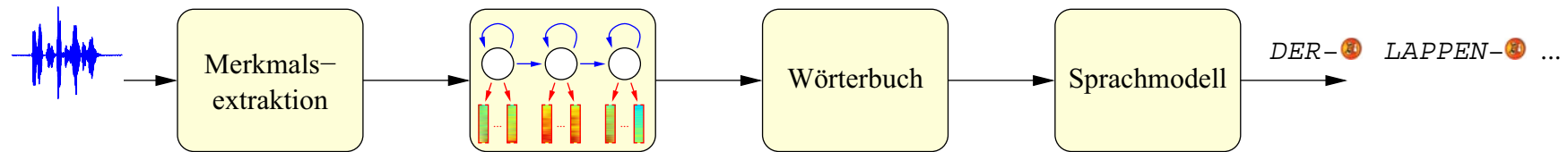
$E(U) = 1$: *Wohin möchten Sie reisen?*

$E(U) = 2$: **Es tut mir Leid**, *dürfte ich bitte erfahren, wohin Sie reisen möchten?*

Sprach-Emotionserkennung



Sprach-Emotionserkennung



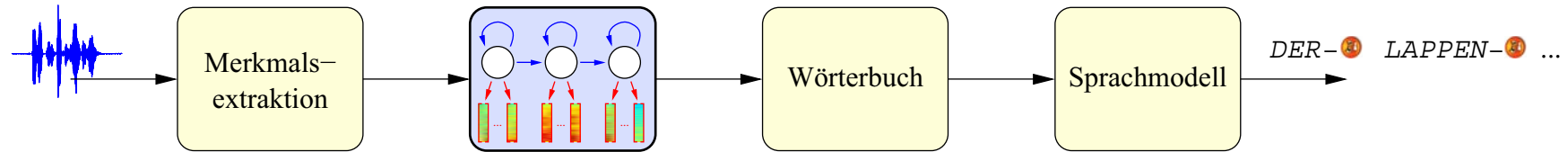
- ▶ HMM – Gaussian Mixtures: $\varphi_j = \sum_{i=0}^M w_{i,j} \cdot \mathcal{N}(\mu_{i,j}, \sigma_{i,j})$

- ▶ Worte \Rightarrow Wort-Emotionen

ABEND	aa	b	aeh	n	t
\Downarrow					
ABEND-FREUDE	aaf	bf	aehf	nf	tf
ABEND-TRAUER	aat	bt	aeht	nt	tt
ABEND-WUT	aaw	bw	aehw	nw	tw

- ▶ Spektrale Merkmale und prosodisch-akustische Merkmale
- ▶ Worterkennerrate $\leq 89\%$, Emotionserkennerrate $\leq 67\%$

Sprach-Emotionserkennung



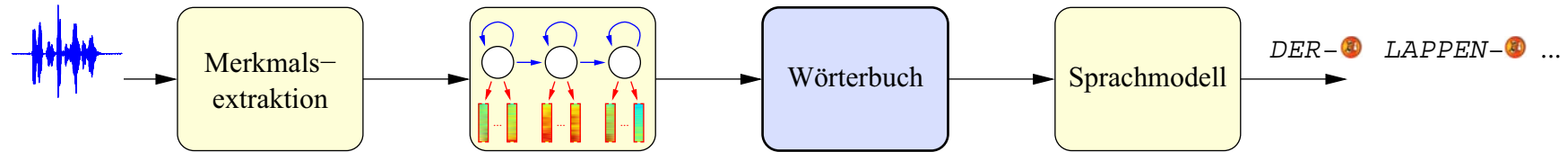
- ▶ HMM – Gaussian Mixtures: $\varphi_j = \sum_{i=0}^M w_{i,j} \cdot \mathcal{N}(\mu_{i,j}, \sigma_{i,j})$

- ▶ Worte \Rightarrow Wort-Emotionen

ABEND		aa	b	aeh	n	t
	↓					
ABEND-FREUDE		aaf	bf	aehf	nf	tf
ABEND-TRAUER		aat	bt	aeht	nt	tt
ABEND-WUT		aaw	bw	aehw	nw	tw

- ▶ Spektrale Merkmale und prosodisch-akustische Merkmale
- ▶ Worterkennerrate $\leq 89\%$, Emotionserkennerrate $\leq 67\%$

Sprach-Emotionserkennung



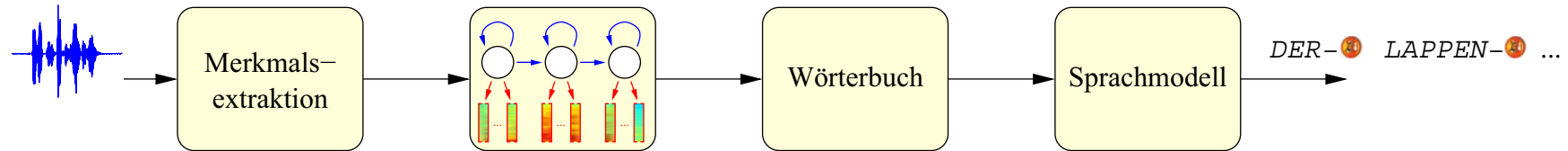
- ▶ HMM – Gaussian Mixtures: $\varphi_j = \sum_{i=0}^M w_{i,j} \cdot \mathcal{N}(\mu_{i,j}, \sigma_{i,j})$

- ▶ Worte \Rightarrow Wort-Emotionen

ABEND	aa	b	aeh	n	t
	⇓				
ABEND-FREUDE	aa ^f	b ^f	aeh ^f	n ^f	t ^f
ABEND-TRAUER	aa ^t	b ^t	aeh ^t	n ^t	t ^t
ABEND-WUT	aa ^w	b ^w	aeh ^w	n ^w	t ^w

- ▶ Spektrale Merkmale und prosodisch-akustische Merkmale
- ▶ Worterkennerrate $\leq 89\%$, Emotionserkennerrate $\leq 67\%$

Sprach-Emotionserkennung



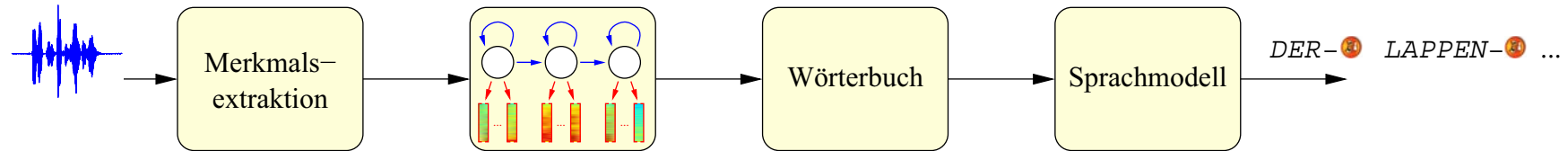
- ▶ HMM – Gaussian Mixtures: $\varphi_j = \sum_{i=0}^M w_{i,j} \cdot \mathcal{N}(\mu_{i,j}, \sigma_{i,j})$

- ▶ Worte \Rightarrow Wort-Emotionen

ABEND	aa	b	aeh	n	t
	⇓				
ABEND-FREUDE	aa f	b f	aeh f	n f	t f
ABEND-TRAUER	aa t	b t	aeh t	n t	t t
ABEND-WUT	aa w	b w	aeh w	n w	t w

- ▶ Spektrale Merkmale und prosodisch-akustische Merkmale
- ▶ Worterkennerrate $\leq 89\%$, Emotionserkennerrate $\leq 67\%$

Sprach-Emotionserkennung



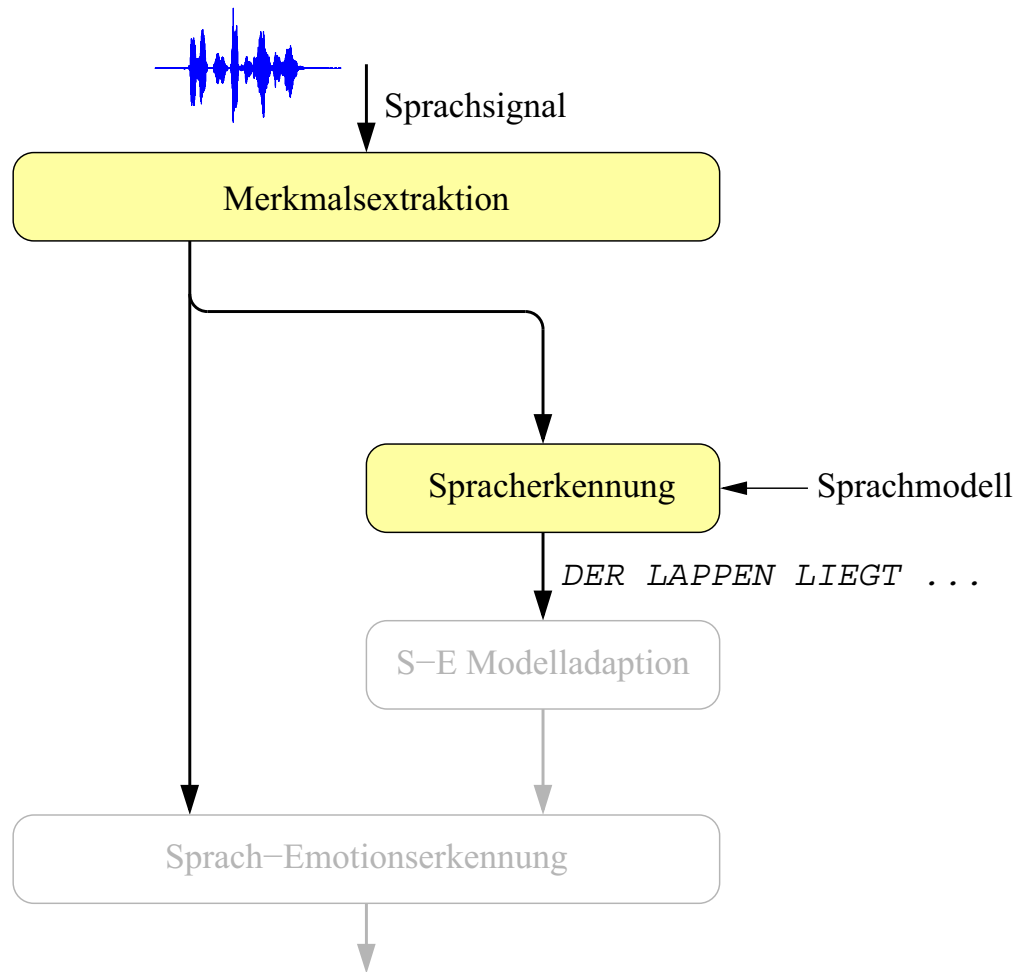
- ▶ HMM – Gaussian Mixtures: $\varphi_j = \sum_{i=0}^M w_{i,j} \cdot \mathcal{N}(\mu_{i,j}, \sigma_{i,j})$

- ▶ Worte \Rightarrow Wort-Emotionen

ABEND	aa	b	aeh	n	t
⇓					
ABEND-FREUDE	aa f	b f	aeh f	n f	t f
ABEND-TRAUER	aa t	b t	aeh t	n t	t t
ABEND-WUT	aa w	b w	aeh w	n w	t w

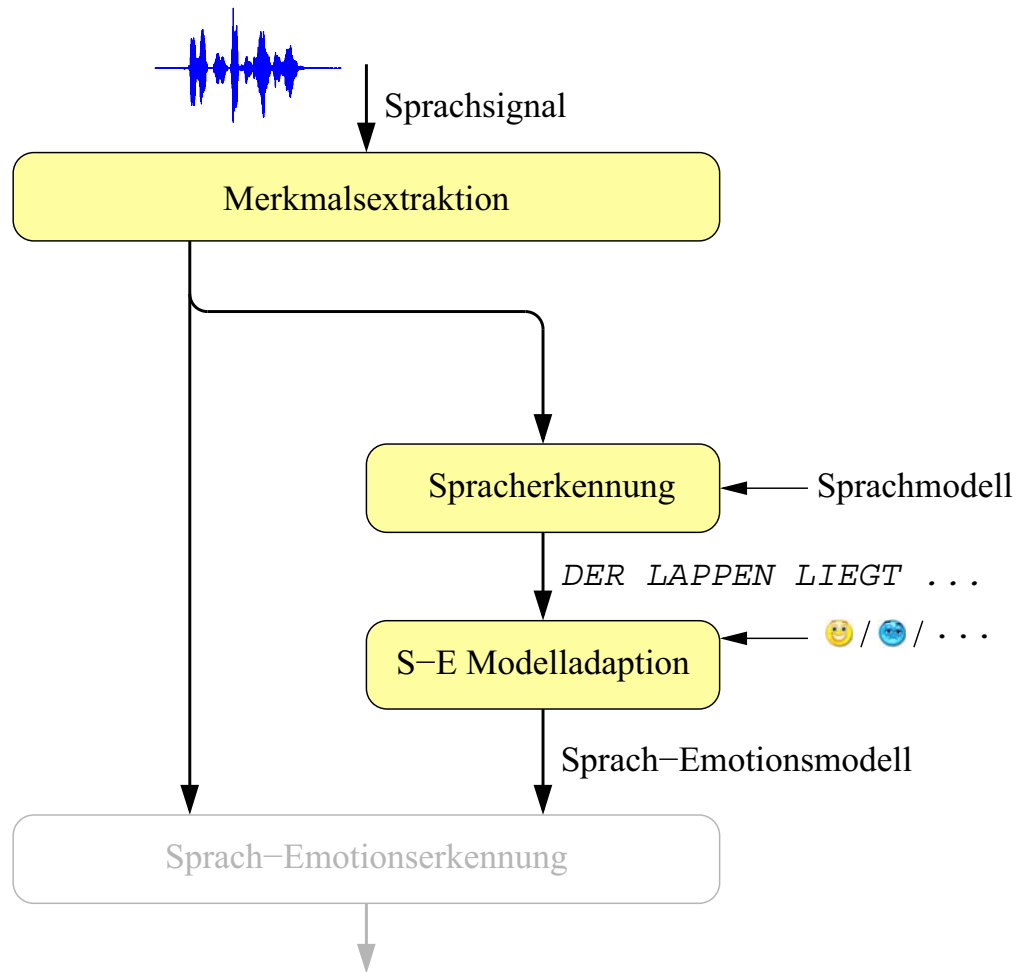
- ▶ Spektrale Merkmale und prosodisch-akustische Merkmale
- ▶ Worterkennerrate $\leq 89\%$, Emotionserkennerrate $\leq 67\%$

Zweistufige Sprach-Emotionserkennung



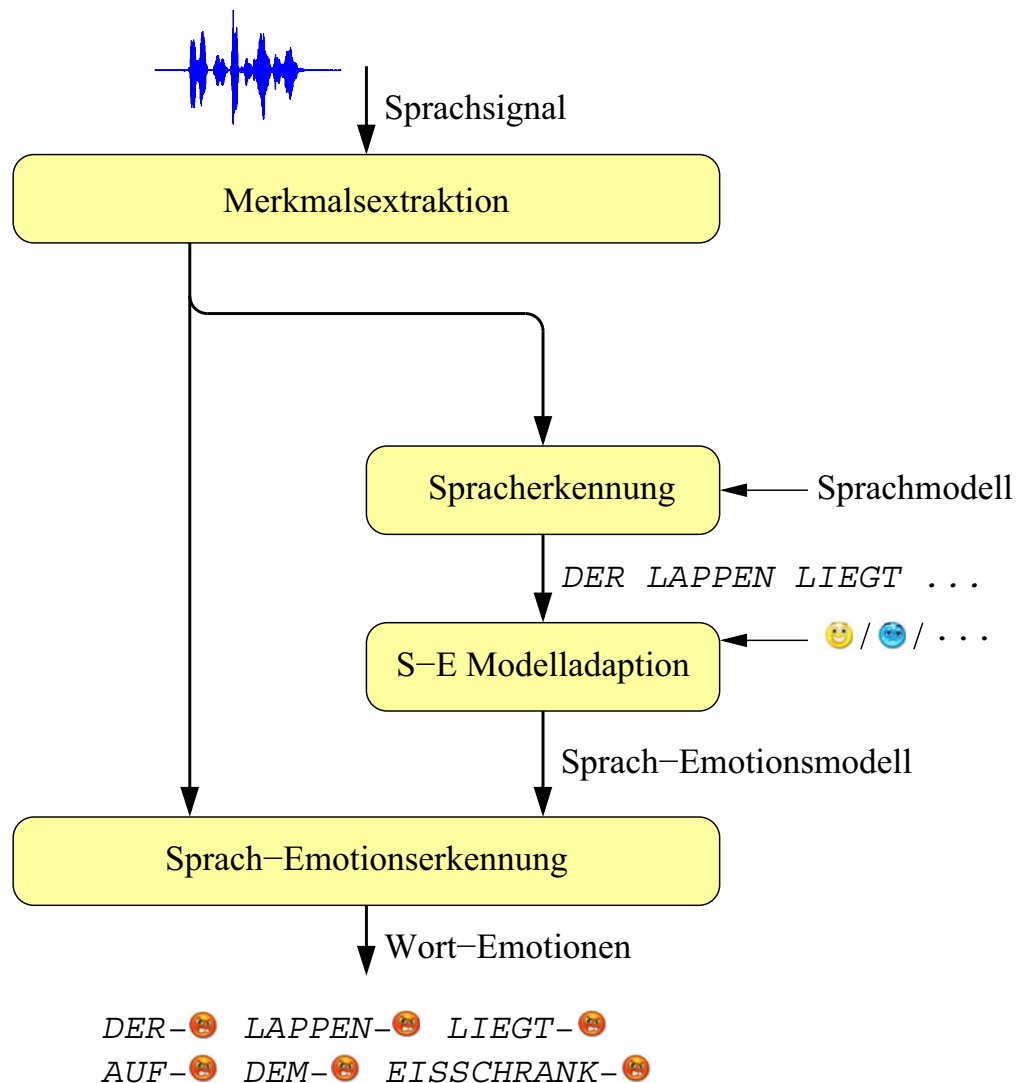
- ▶ Text als a-priori-Information für Sprach-Emotionserkennung
- ▶ Worterkennerrate **94.7%**
- ▶ Emotionserkennerrate **72.6%**

Zweistufige Sprach-Emotionserkennung



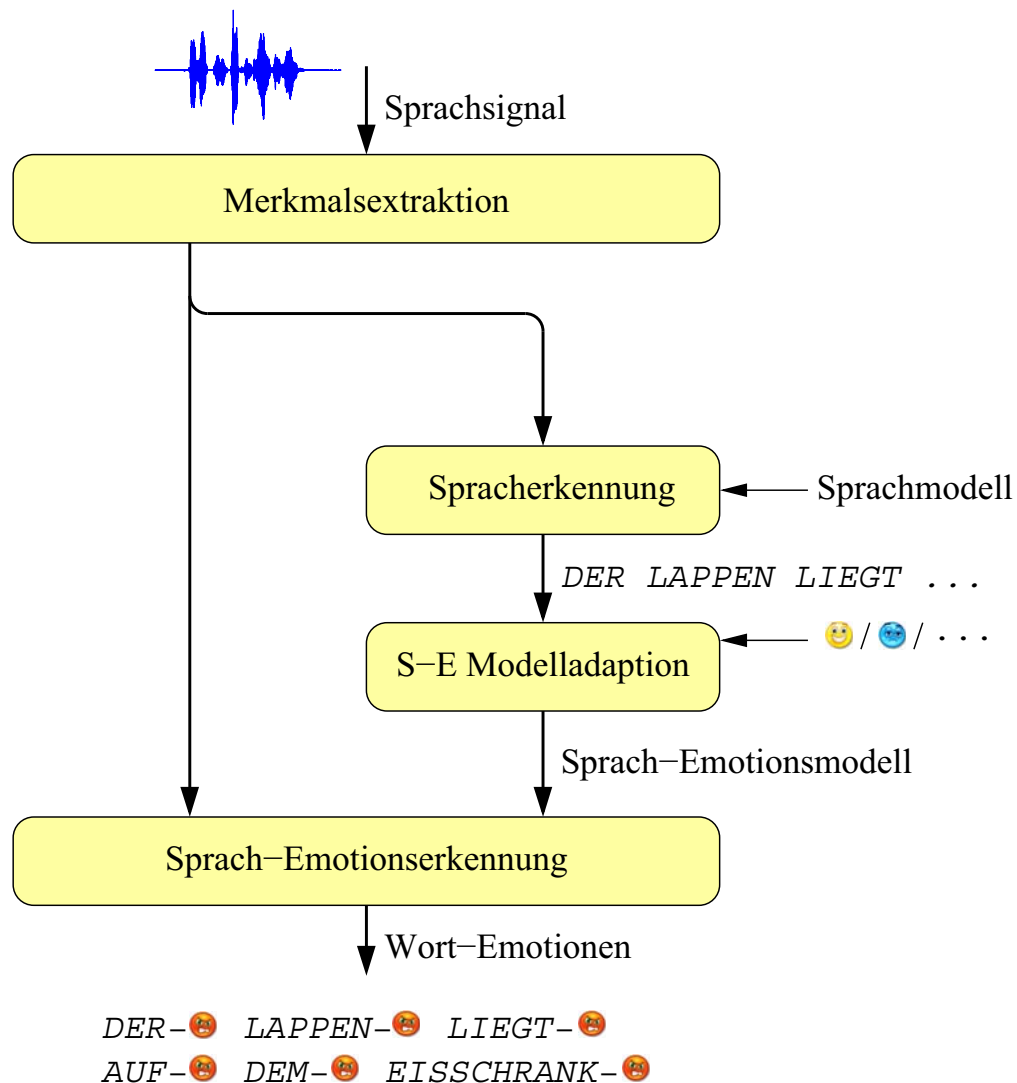
- ▶ Text als a-priori-Information für Sprach-Emotionserkennung
- ▶ Worterkennerrate **94.7%**
- ▶ Emotionserkennerrate **72.6%**

Zweistufige Sprach-Emotionserkennung



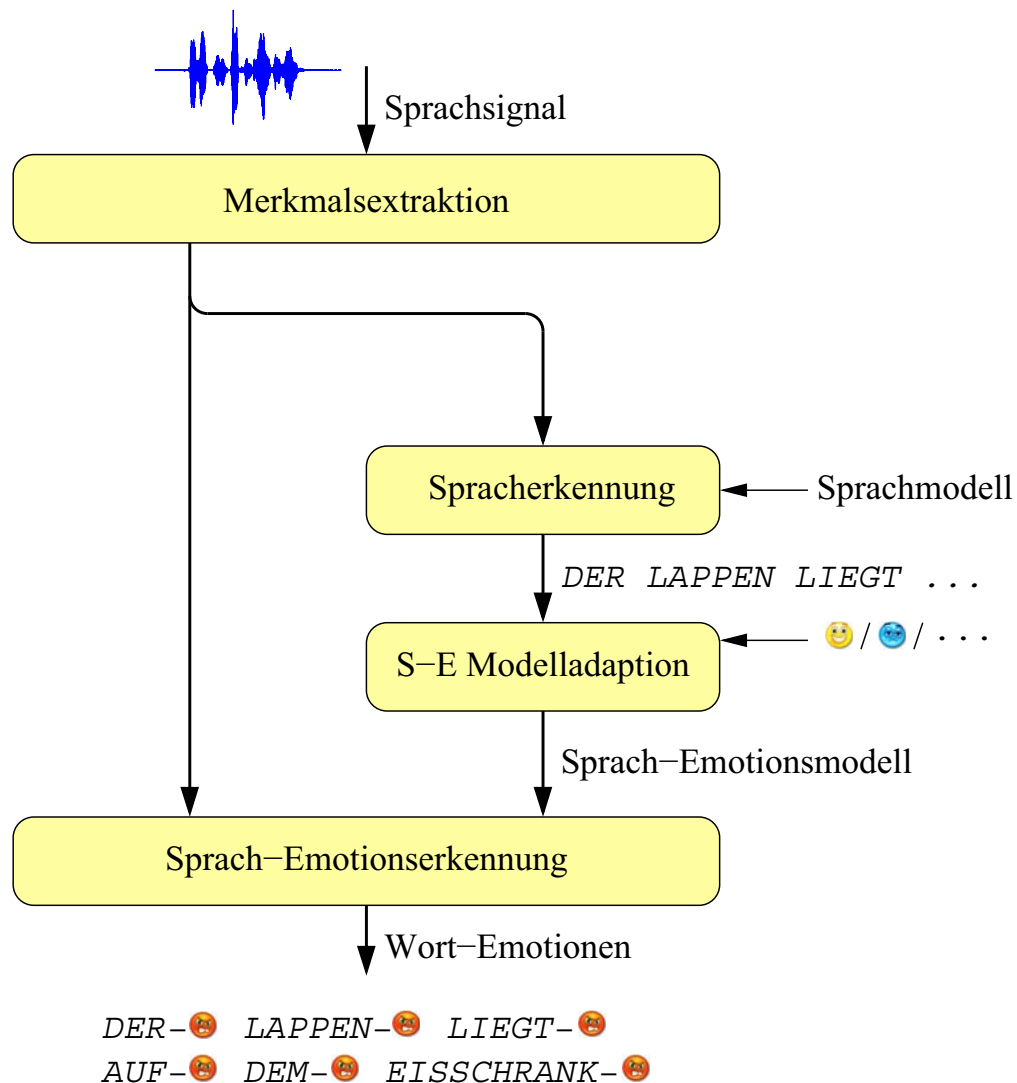
- ▶ Text als a-priori-Information für Sprach-Emotionserkennung
- ▶ Worterkennerrate **94.7%**
- ▶ Emotionserkennerrate **72.6%**

Zweistufige Sprach-Emotionserkennung



- ▶ Text als a-priori-Information für Sprach-Emotionserkennung
- ▶ Worterkennerrate **94.7%**
- ▶ Emotionserkennerrate **72.6%**

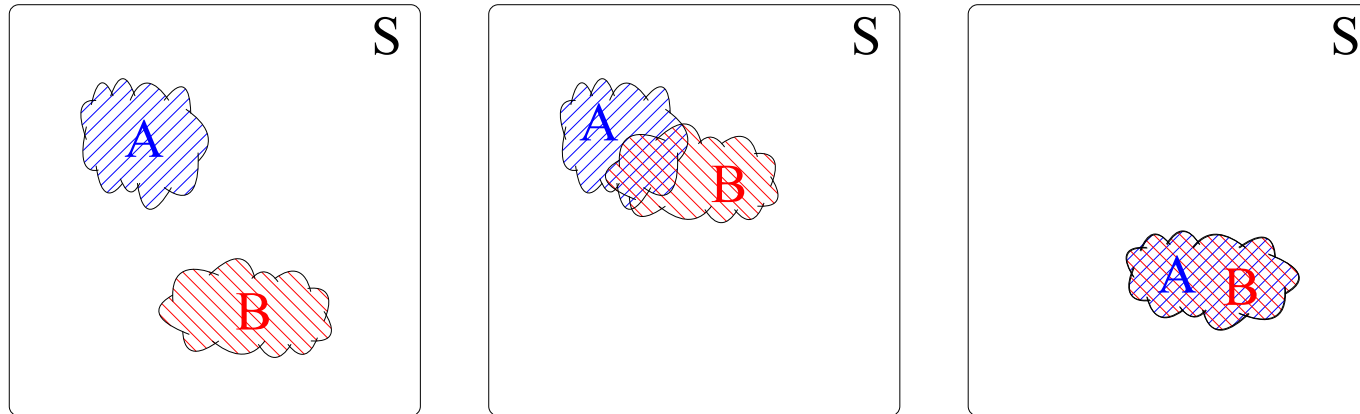
Zweistufige Sprach-Emotionserkennung



- ▶ Text als a-priori-Information für Sprach-Emotionserkennung
- ▶ Worterkennerrate **94.7%**
- ▶ Emotionserkennerrate **72.6%**

Mehrere Sprach-Emotionserkenner

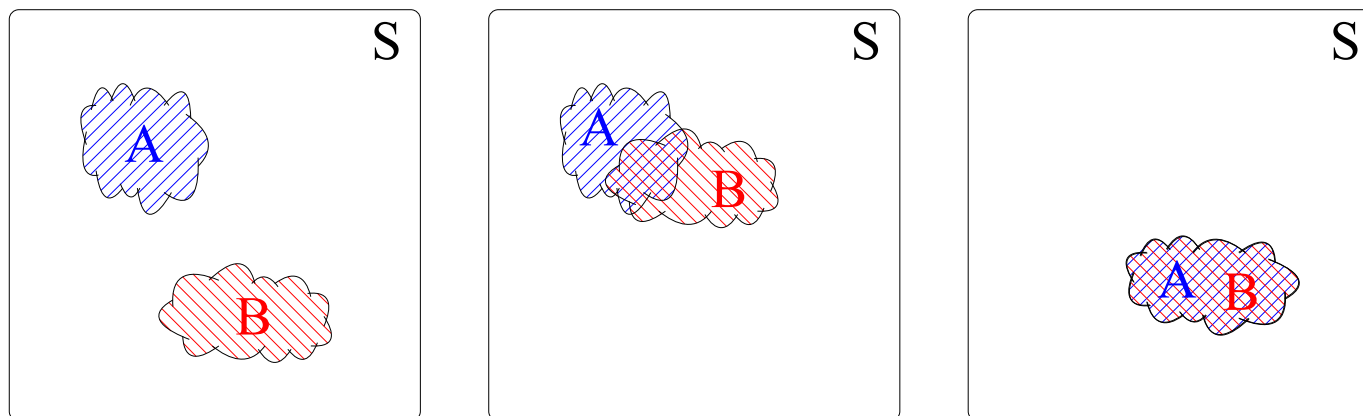
- ▶ Unterschiede in Fehlercharakteristika begünstigen Fehlerkorrekturen



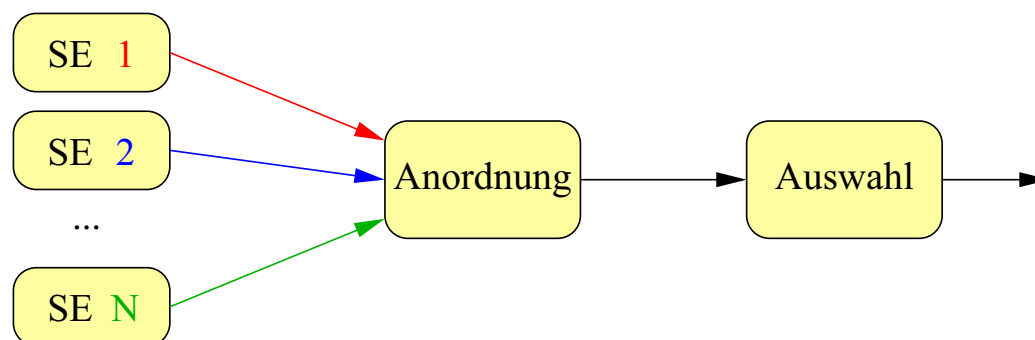
- ▶ ROVER (Recognizer Output Voting Error Reduction) für Spracherkenner (SE)

Mehrere Sprach-Emotionserkenner

- ▶ Unterschiede in Fehlercharakteristika begünstigen Fehlerkorrekturen

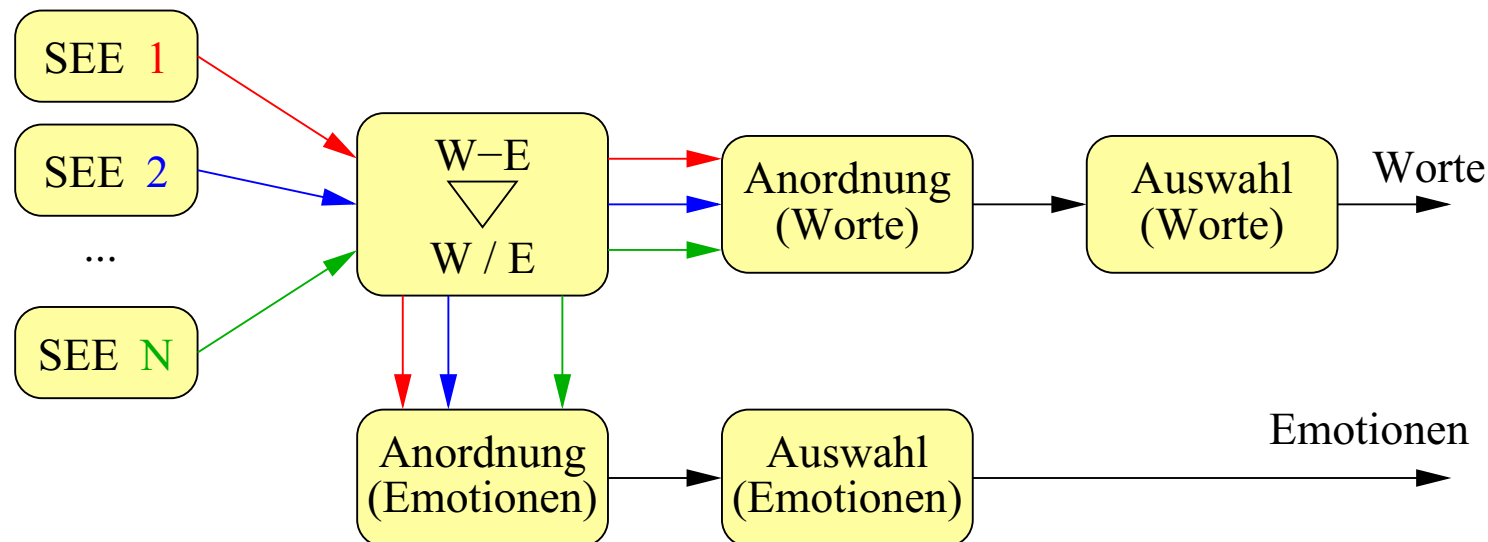


- ▶ ROVER (Recognizer Output Voting Error Reduction) für Spracherkenner (SE)



Mehrere Sprach-Emotionserkenner

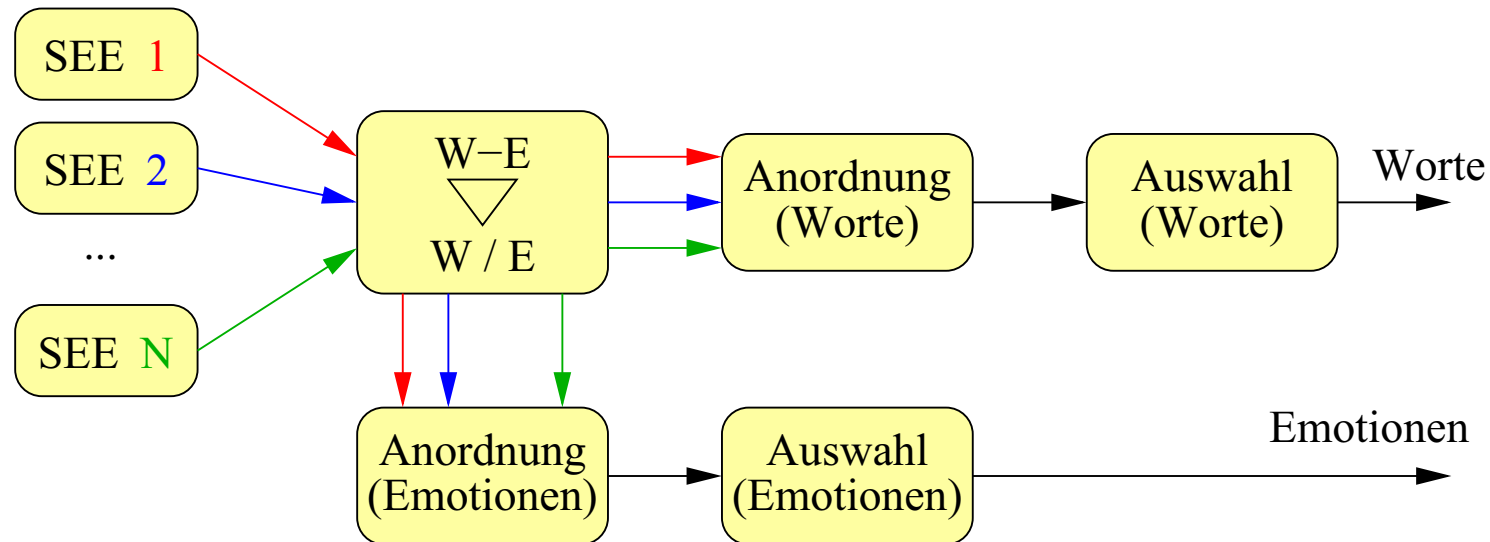
- ▶ ROVER für Sprach-Emotionserkenner (SEE)



- ▶ $S(w, e) = \alpha \cdot N'(w, e) + (1 - \alpha) \cdot C(w, e)$
- ▶ Kombination von 2-5 Sprach-Emotionserkennern
- ▶ Worterkennerraten [82.5%, 89.1%] \Rightarrow **89.8%**
- ▶ Emotionserkennerraten [62.3% 68.9%] \Rightarrow **76.4%**

Mehrere Sprach-Emotionserkenner

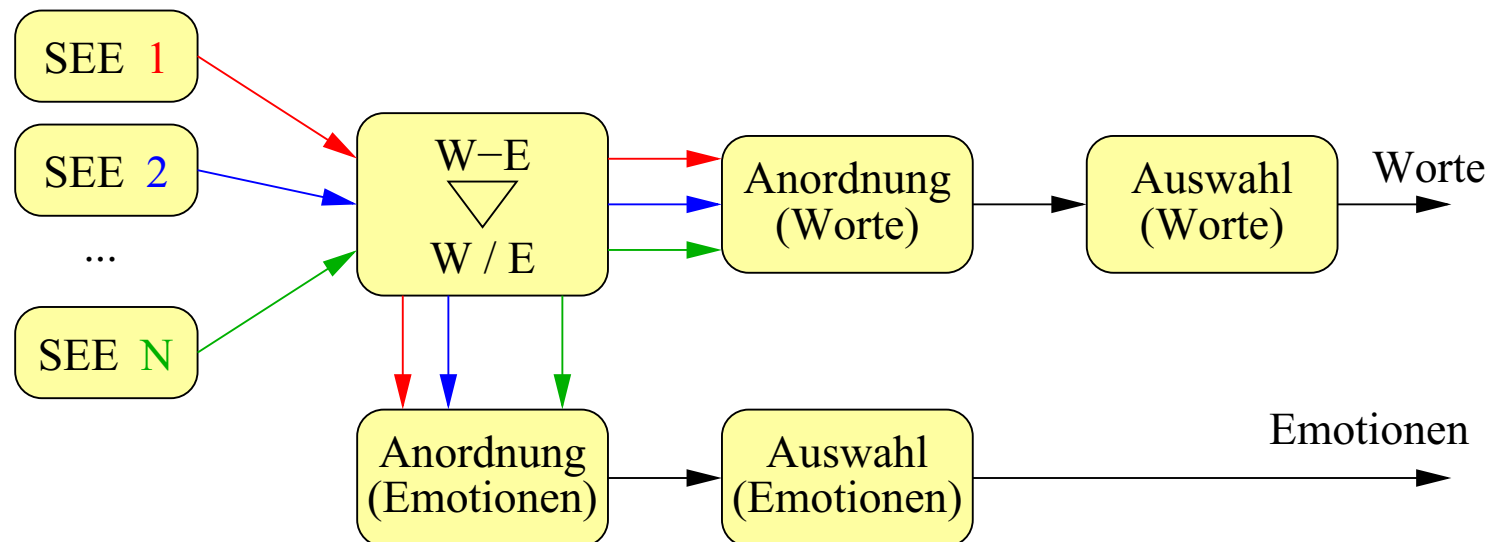
- ▶ ROVER für Sprach-Emotionserkenner (SEE)



- ▶ $S(w, e) = \alpha \cdot N'(w, e) + (1 - \alpha) \cdot C(w, e)$
- ▶ Kombination von 2-5 Sprach-Emotionserkennern
- ▶ Worterkennerraten [82.5%, 89.1%] \Rightarrow **89.8%**
- ▶ Emotionserkennerraten [62.3% 68.9%] \Rightarrow **76.4%**

Mehrere Sprach-Emotionserkenner

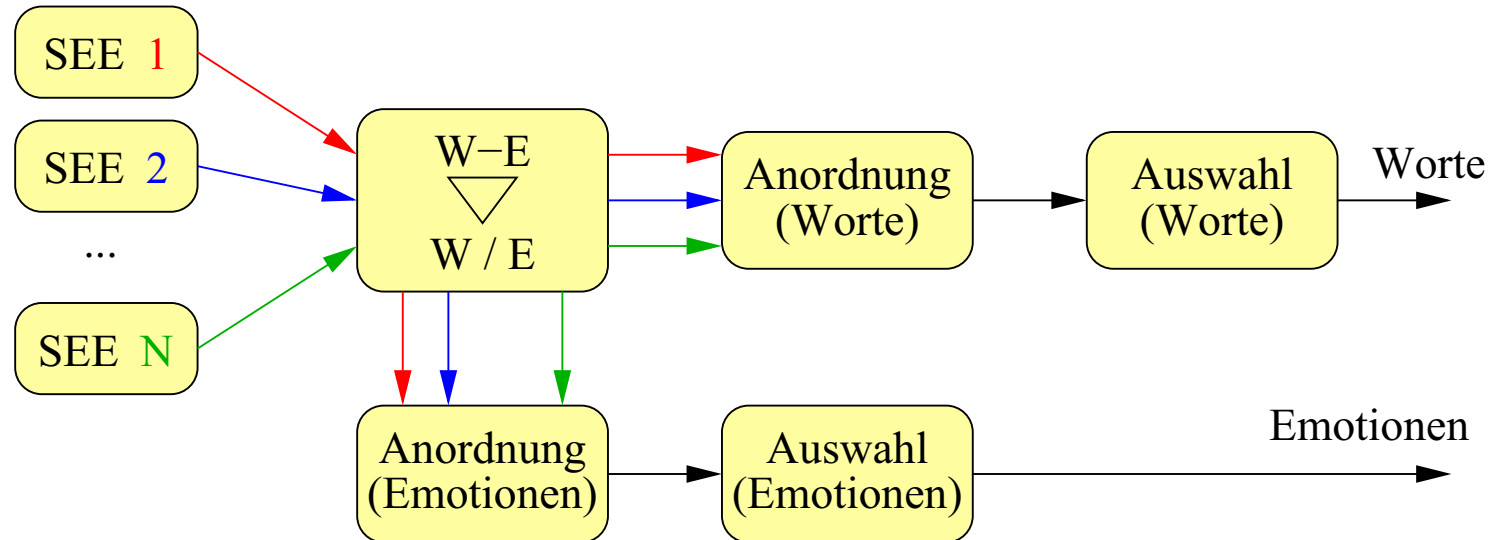
- ▶ ROVER für Sprach-Emotionserkenner (SEE)



- ▶ $S(w, e) = \alpha \cdot N'(w, e) + (1 - \alpha) \cdot C(w, e)$
- ▶ Kombination von 2-5 Sprach-Emotionserkennern
- ▶ Worterkennerraten [82.5%, 89.1%] \Rightarrow **89.8%**
- ▶ Emotionserkennerraten [62.3% 68.9%] \Rightarrow **76.4%**

Mehrere Sprach-Emotionserkenner

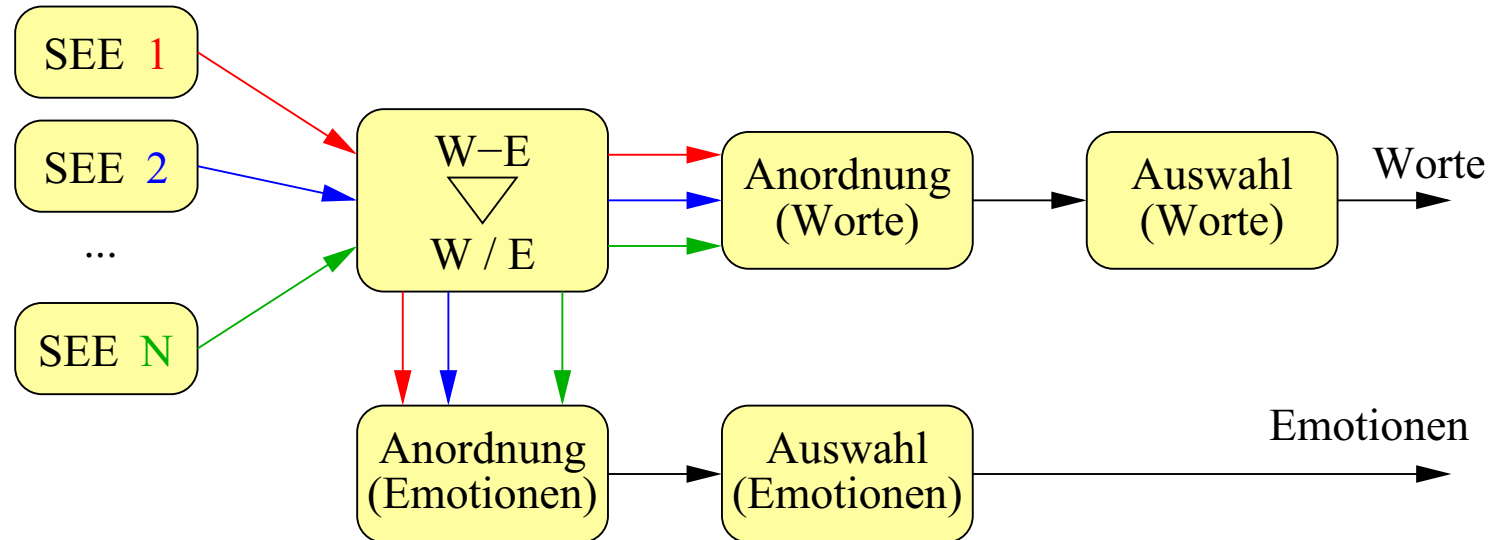
- ▶ ROVER für Sprach-Emotionserkenner (SEE)



- ▶ $S(w, e) = \alpha \cdot N'(w, e) + (1 - \alpha) \cdot C(w, e)$
- ▶ Kombination von 2-5 Sprach-Emotionserkennern
- ▶ Worterkennerraten [82.5%, 89.1%] \Rightarrow **89.8%**
- ▶ Emotionserkennerraten [62.3% 68.9%] \Rightarrow **76.4%**

Mehrere Sprach-Emotionserkenner

- ▶ ROVER für Sprach-Emotionserkenner (SEE)



- ▶ $S(w, e) = \alpha \cdot N'(w, e) + (1 - \alpha) \cdot C(w, e)$
- ▶ Kombination von 2-5 Sprach-Emotionserkennern
- ▶ Worterkennerraten [82.5%, 89.1%] \Rightarrow **89.8%**
- ▶ Emotionserkennerraten [62.3% 68.9%] \Rightarrow **76.4%**

Semantische Analyse

- ▶ Interpretation relevanter Wörter (\rightsquigarrow BEEV)

- ▶ Emotionales Wörterbuch (376 Einträge)

akzeptabel	Neutral	0
<i>[bel. Schimpfwort]</i>	Wut	--
ok	Neutral	0
pfui	Ekel	-
schade	Trauer	-
schön	Freude	++
toll	Freude	++
übel	Ekel	-
wow	Freude	++

- ▶ Integration in Anwendungsgrammatik

- ▶ Emotionserkennerrate ca. **60%**

Semantische Analyse

- ▶ Interpretation relevanter Wörter (\rightsquigarrow BEEV)
- ▶ Emotionales Wörterbuch (376 Einträge)

akzeptabel	Neutral	0
<i>[bel. Schimpfwort]</i>	Wut	--
ok	Neutral	0
pfui	Ekel	-
schade	Trauer	-
schön	Freude	++
toll	Freude	++
übel	Ekel	-
wow	Freude	++

- ▶ Integration in Anwendungsgrammatik
- ▶ Emotionserkennerrate ca. 60%

Semantische Analyse

- ▶ Interpretation relevanter Wörter (\rightsquigarrow BEEV)
- ▶ Emotionales Wörterbuch (376 Einträge)

akzeptabel	Neutral	0
<i>[bel. Schimpfwort]</i>	Wut	--
ok	Neutral	0
pfui	Ekel	-
schade	Trauer	-
schön	Freude	++
toll	Freude	++
übel	Ekel	-
wow	Freude	++

- ▶ Integration in Anwendungsgrammatik
- ▶ Emotionserkennerrate ca. 60%

Semantische Analyse

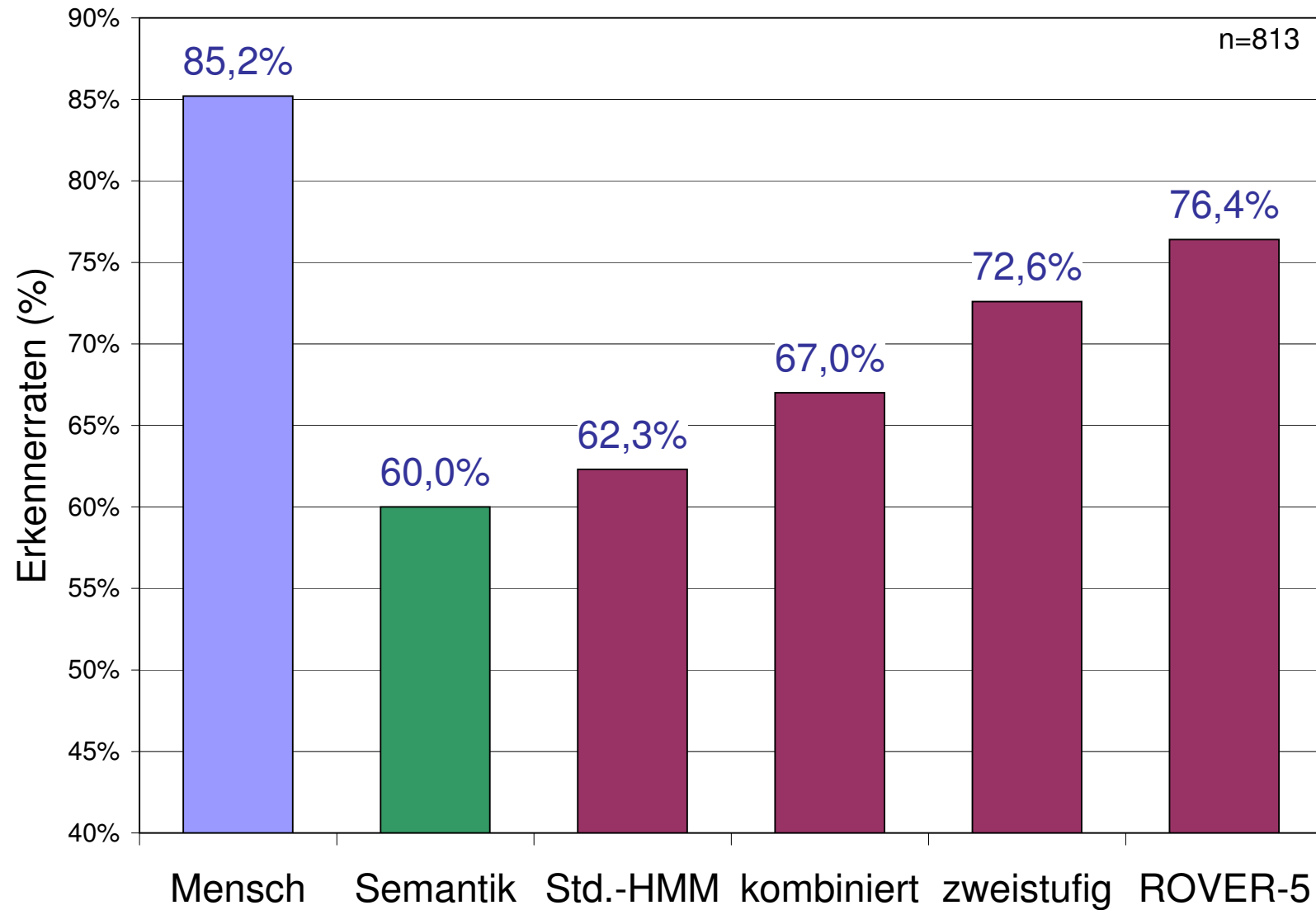
- ▶ Interpretation relevanter Wörter (\rightsquigarrow BEEV)

- ▶ Emotionales Wörterbuch (376 Einträge)

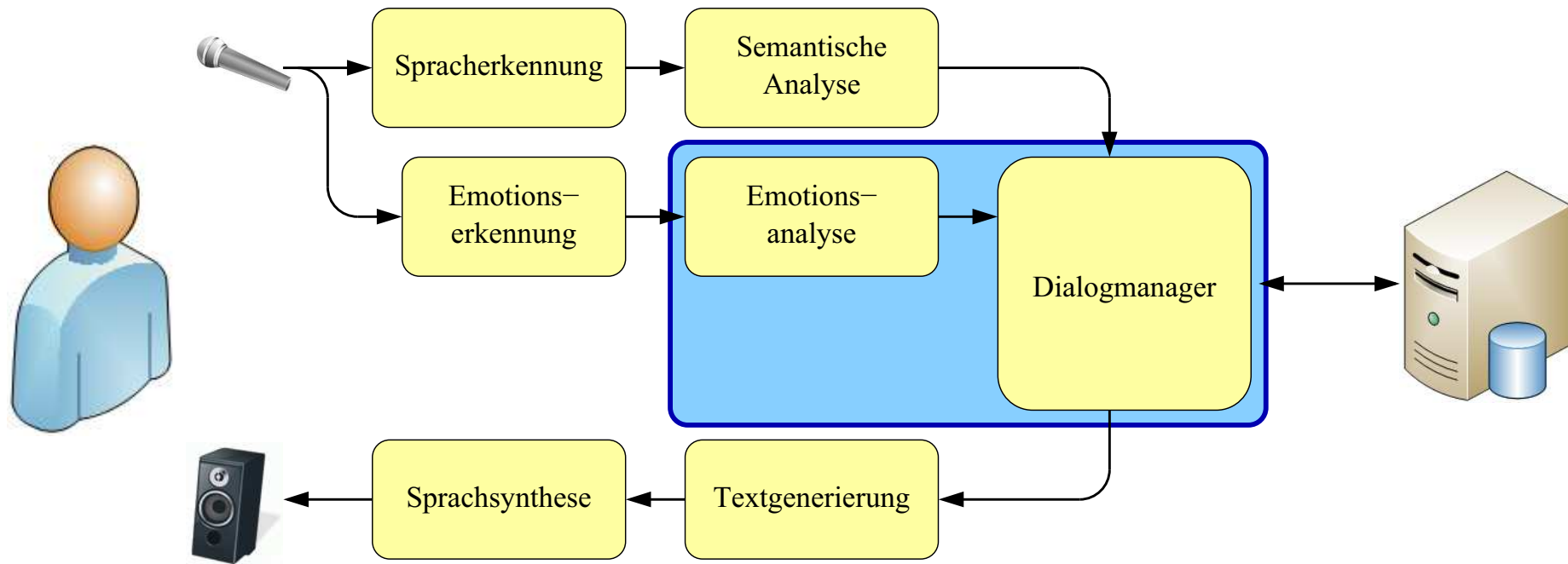
akzeptabel	Neutral	0
<i>[bel. Schimpfwort]</i>	Wut	--
ok	Neutral	0
pfui	Ekel	-
schade	Trauer	-
schön	Freude	++
toll	Freude	++
übel	Ekel	-
wow	Freude	++

- ▶ Integration in Anwendungsgrammatik
- ▶ Emotionserkennerrate ca. **60%**

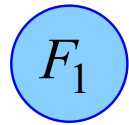
Emotionserkennerraten



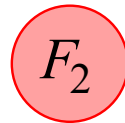
Emotionen im Dialogmanagement



Dialogmodell



F_1



F_2

- ▶ Zustände und Funktion vorgegeben

- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten



F_4



F_3

- ▶ $P(F_i|F_j)$ ("bi-turns")
- ▶ $P(F_i|F_j, F_k)$ ("tri-turns")

- ▶ Training mit aufbereiteten Dialogdaten

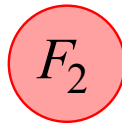
Zielort → Abfahrtsort

Zielort, Datum → Zeit

Dialogmodell



F_1



F_2

- ▶ Zustände und Funktion vorgegeben



F_4



F_3

- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten

- ▶ $P(F_i|F_j)$ (“bi-turns”)

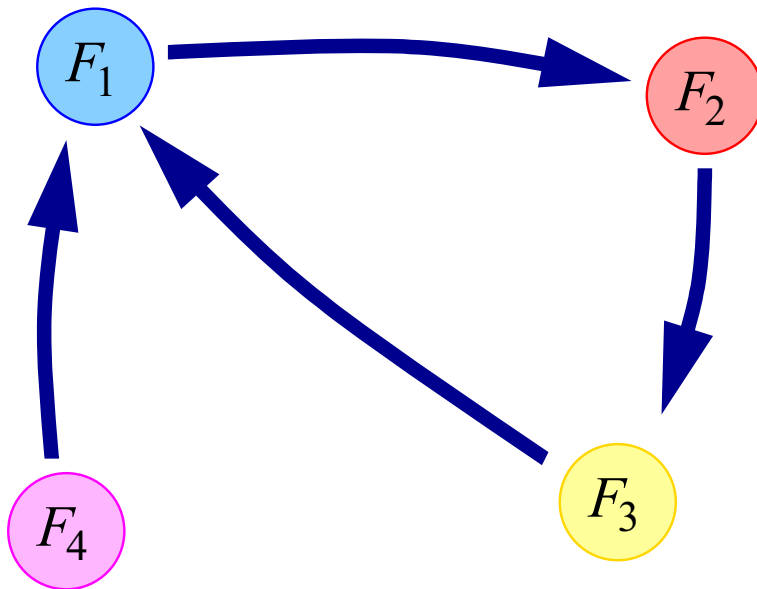
- ▶ $P(F_i|F_j, F_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Zielort → Abfahrtsort

Zielort, Datum → Zeit

Dialogmodell



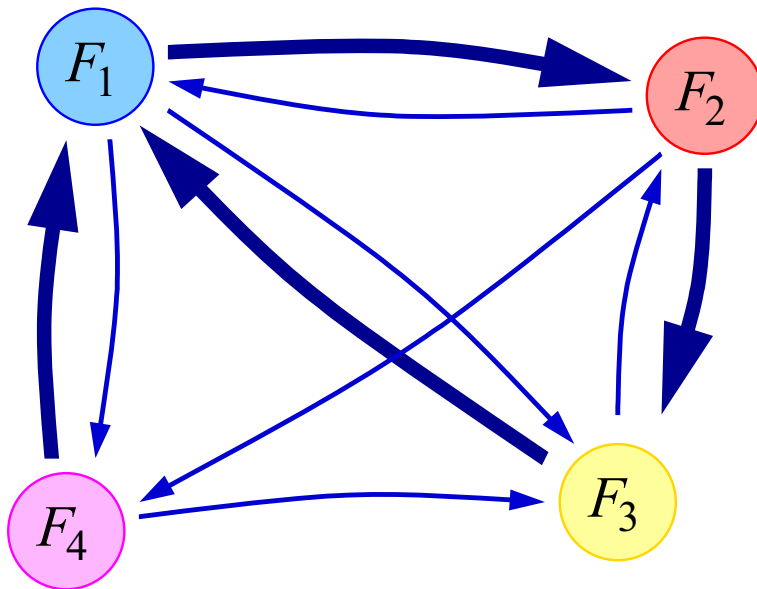
- ▶ Zustände und Funktion vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(F_i|F_j)$ (“bi-turns”)
 - ▶ $P(F_i|F_j, F_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Zielort → Abfahrtsort

Zielort, Datum → Zeit

Dialogmodell



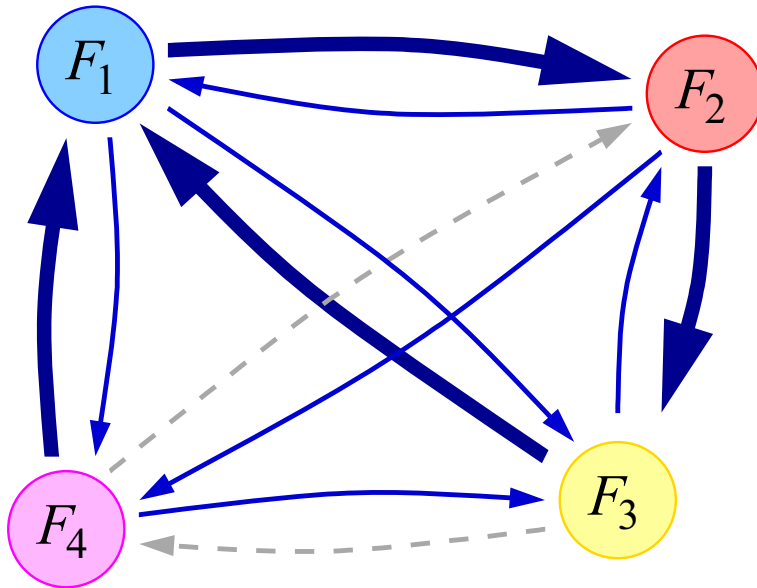
- ▶ Zustände und Funktion vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(F_i|F_j)$ (“bi-turns”)
 - ▶ $P(F_i|F_j, F_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Zielort → Abfahrtsort

Zielort, Datum → Zeit

Dialogmodell



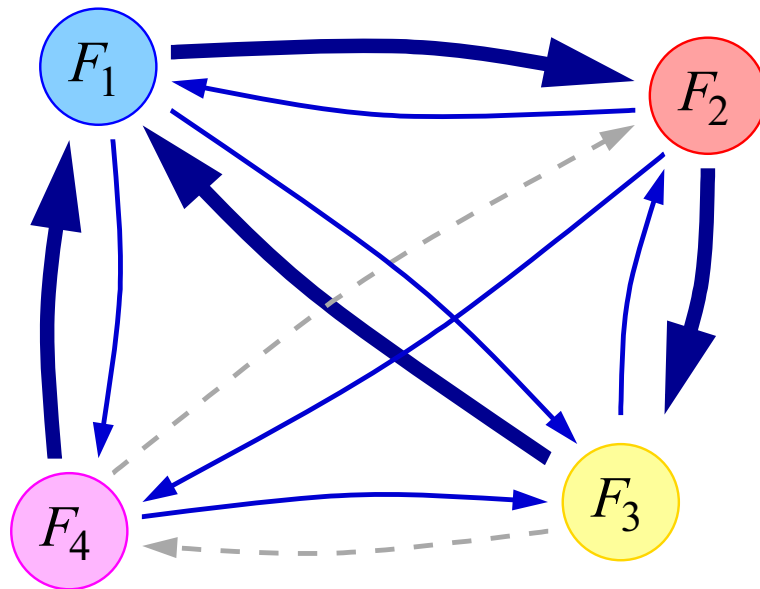
- ▶ Zustände und Funktion vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(F_i|F_j)$ (“bi-turns”)
 - ▶ $P(F_i|F_j, F_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Zielort → Abfahrtsort

Zielort, Datum → Zeit

Dialogmodell



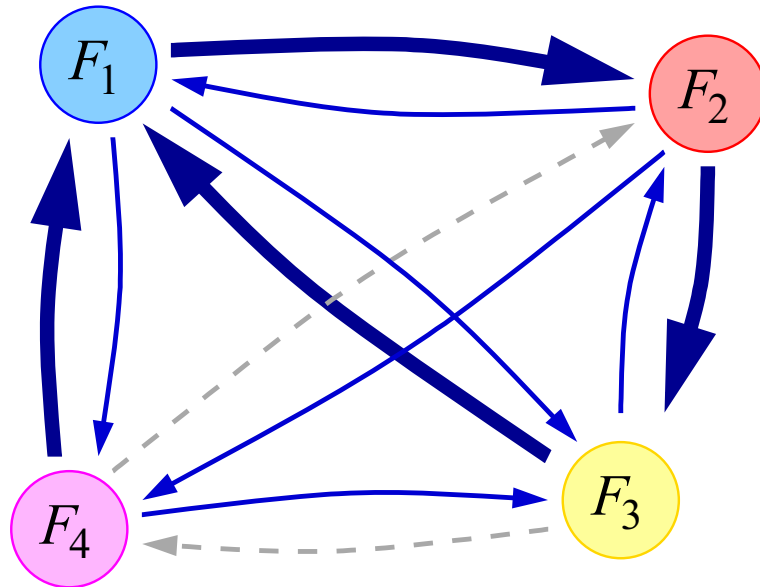
- ▶ Zustände und Funktion vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(F_i|F_j)$ (“bi-turns”)
 - ▶ $P(F_i|F_j, F_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Zielort → Abfahrtsort

Zielort, Datum → Zeit

Dialogmodell



- ▶ Zustände und Funktion vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(F_i|F_j)$ (“bi-turns”)
 - ▶ $P(F_i|F_j, F_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Zielort → Abfahrtsort

Zielort, Datum → Zeit

Emotionsmodell

E_1

E_2

E_3

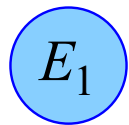
- ▶ Emotionale Zustände vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(E_i|E_j)$ (“bi-turns”)
 - ▶ $P(E_i|E_j, E_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Freude → Wut

Trauer, Freude → Freude

Emotionsmodell



E_1



E_2



E_3

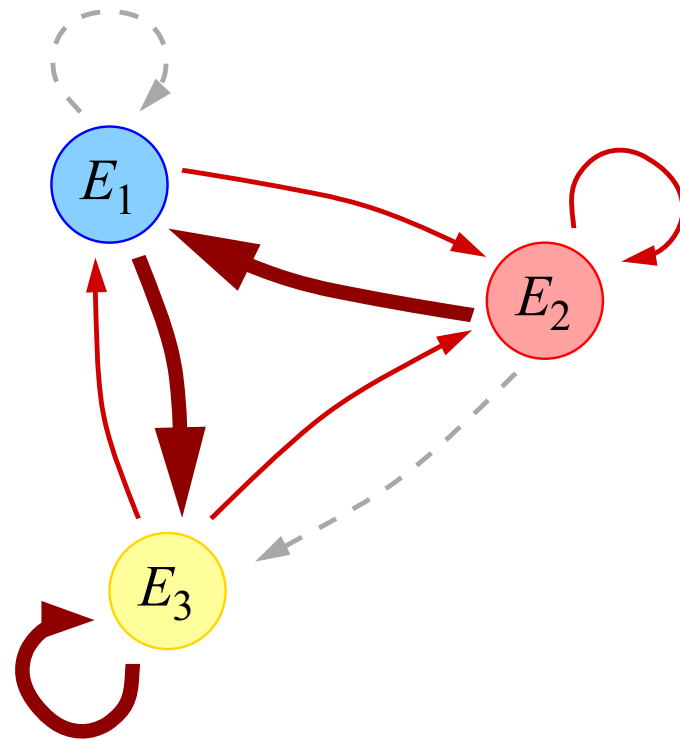
- ▶ Emotionale Zustände vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(E_i|E_j)$ (“bi-turns”)
 - ▶ $P(E_i|E_j, E_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Freude → Wut

Trauer, Freude → Freude

Emotionsmodell



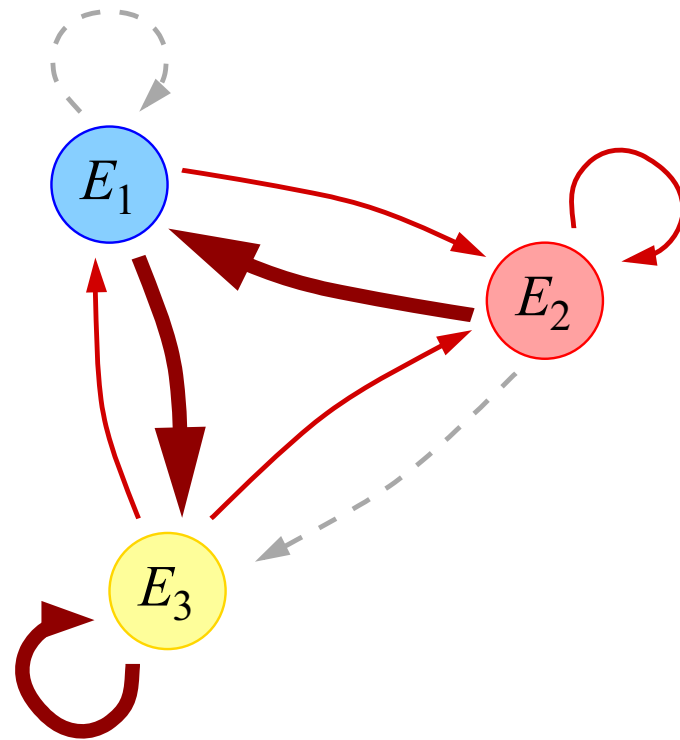
- ▶ Emotionale Zustände vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(E_i|E_j)$ (“bi-turns”)
 - ▶ $P(E_i|E_j, E_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Freude → Wut

Trauer, Freude → Freude

Emotionsmodell



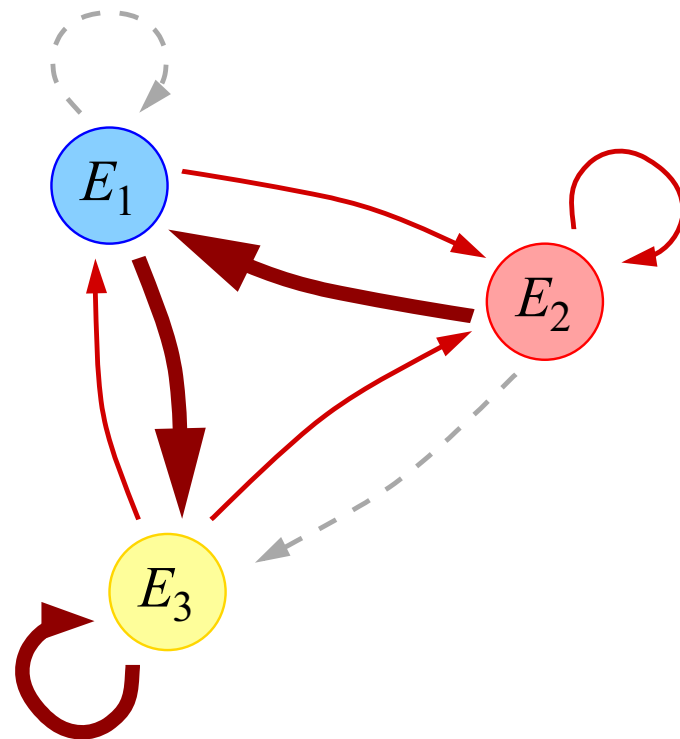
- ▶ Emotionale Zustände vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(E_i|E_j)$ (“bi-turns”)
 - ▶ $P(E_i|E_j, E_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Freude → Wut

Trauer, Freude → Freude

Emotionsmodell



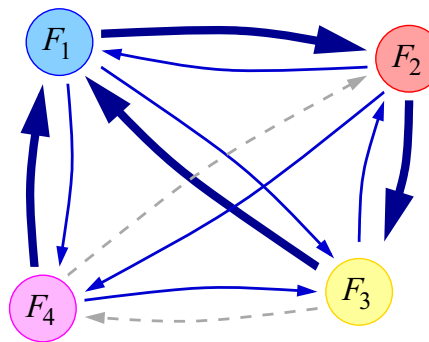
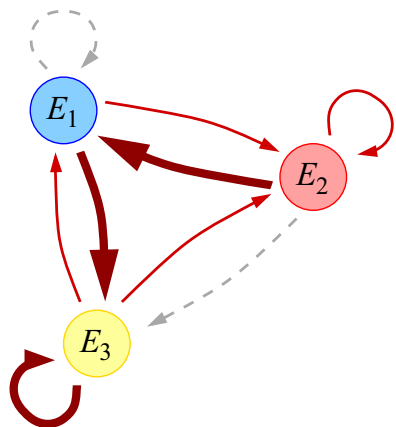
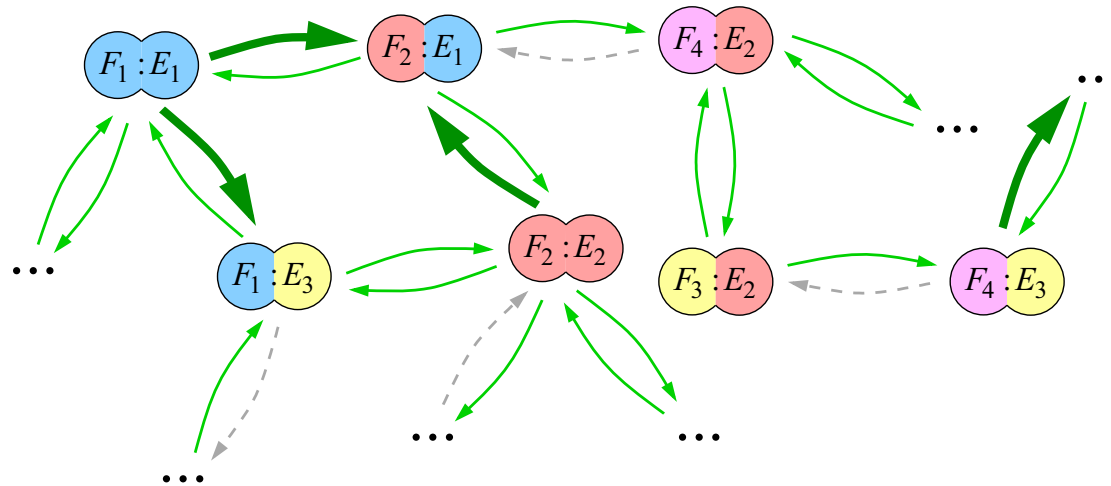
- ▶ Emotionale Zustände vorgegeben
- ▶ Übergänge zwischen Zuständen definiert durch Wahrscheinlichkeiten
 - ▶ $P(E_i|E_j)$ (“bi-turns”)
 - ▶ $P(E_i|E_j, E_k)$ (“tri-turns”)

- ▶ Training mit aufbereiteten Dialogdaten

Freude \rightarrow Wut

Trauer, Freude \rightarrow Freude

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte."

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin."

S: "Es tut mir leid. Wann möchten Sie abreisen?"

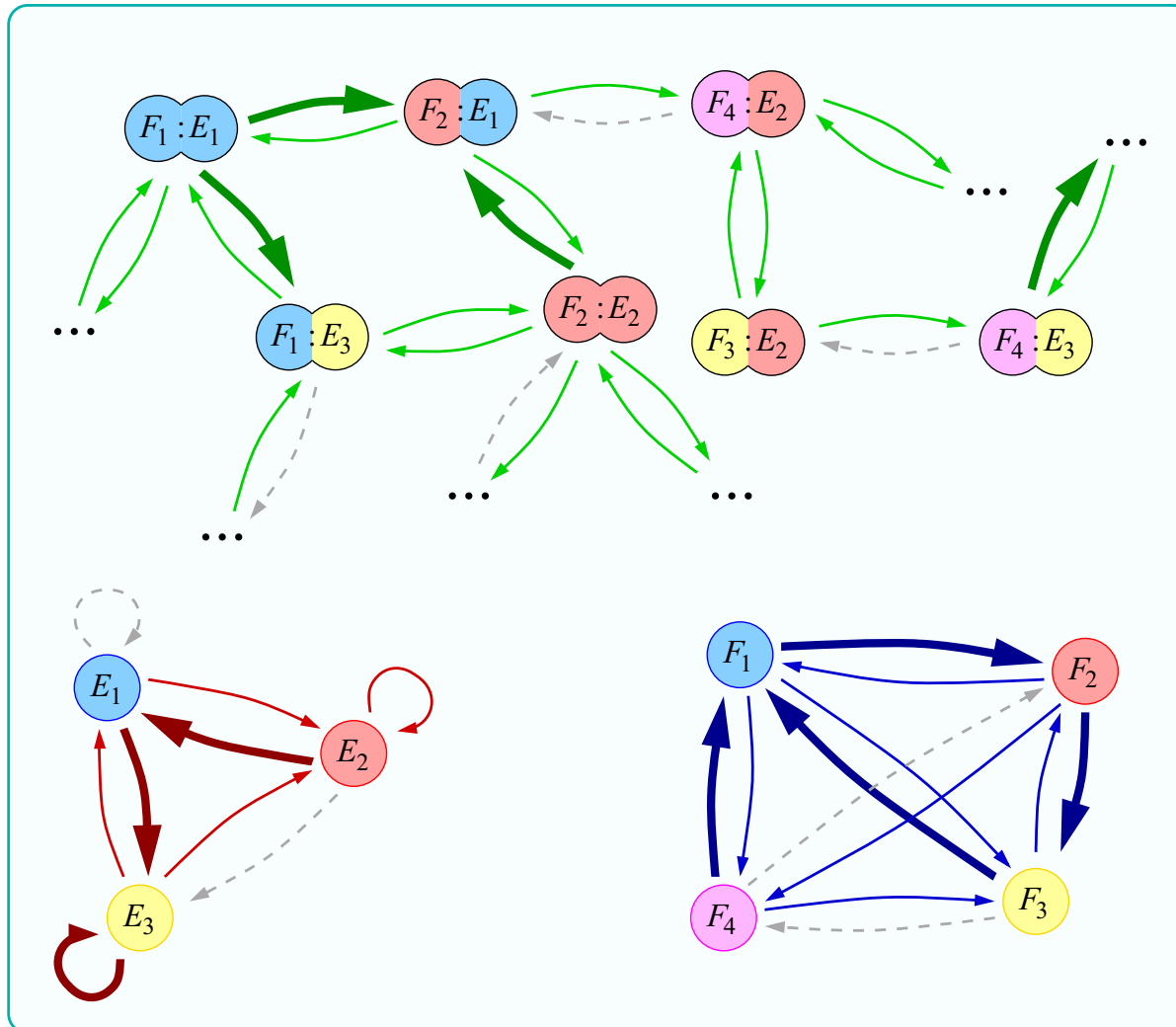
B: "Um 17 Uhr."

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte."

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin."

S: "Es tut mir leid. Wann möchten Sie abreisen?"

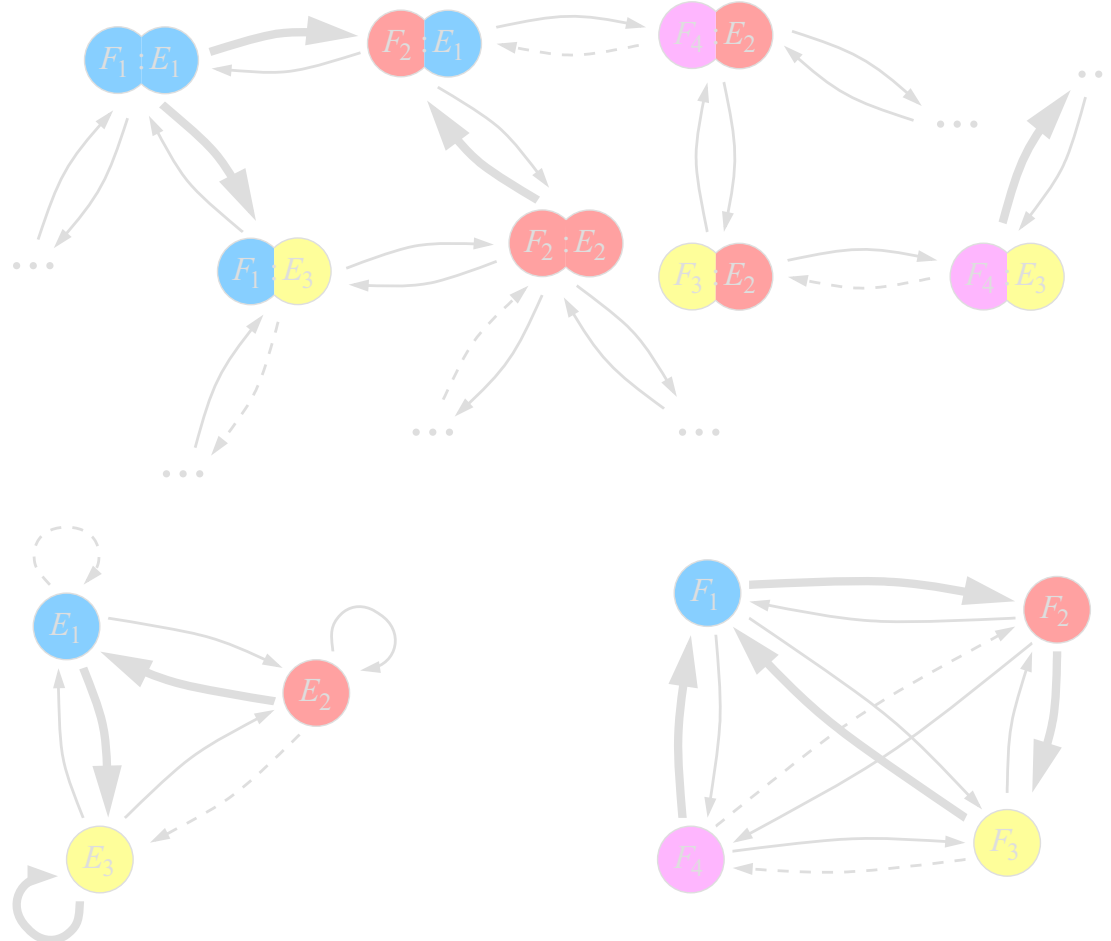
B: "Um 17 Uhr."

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte."

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin."

S: "Es tut mir leid. Wann möchten Sie abreisen?"

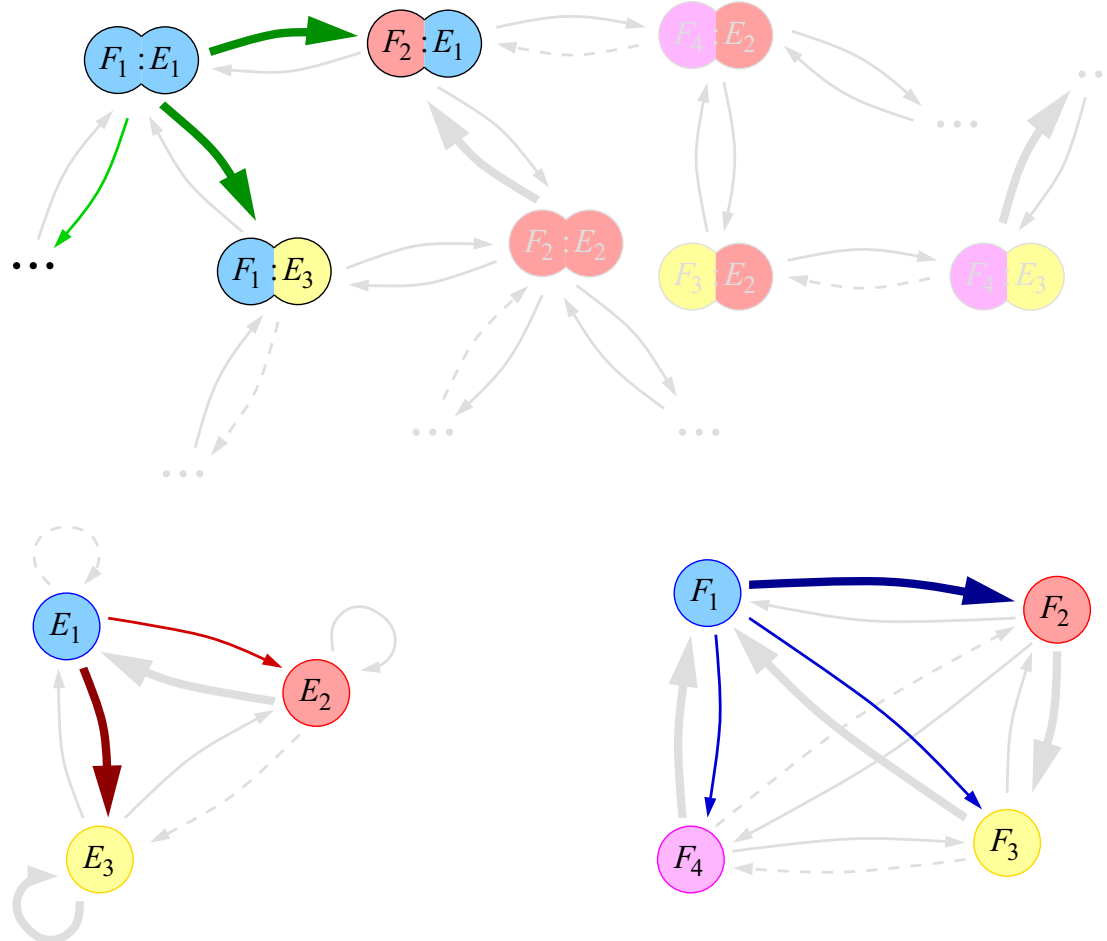
B: "Um 17 Uhr."

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte." 🤖

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin."

S: "Es tut mir leid. Wann möchten Sie abreisen?"

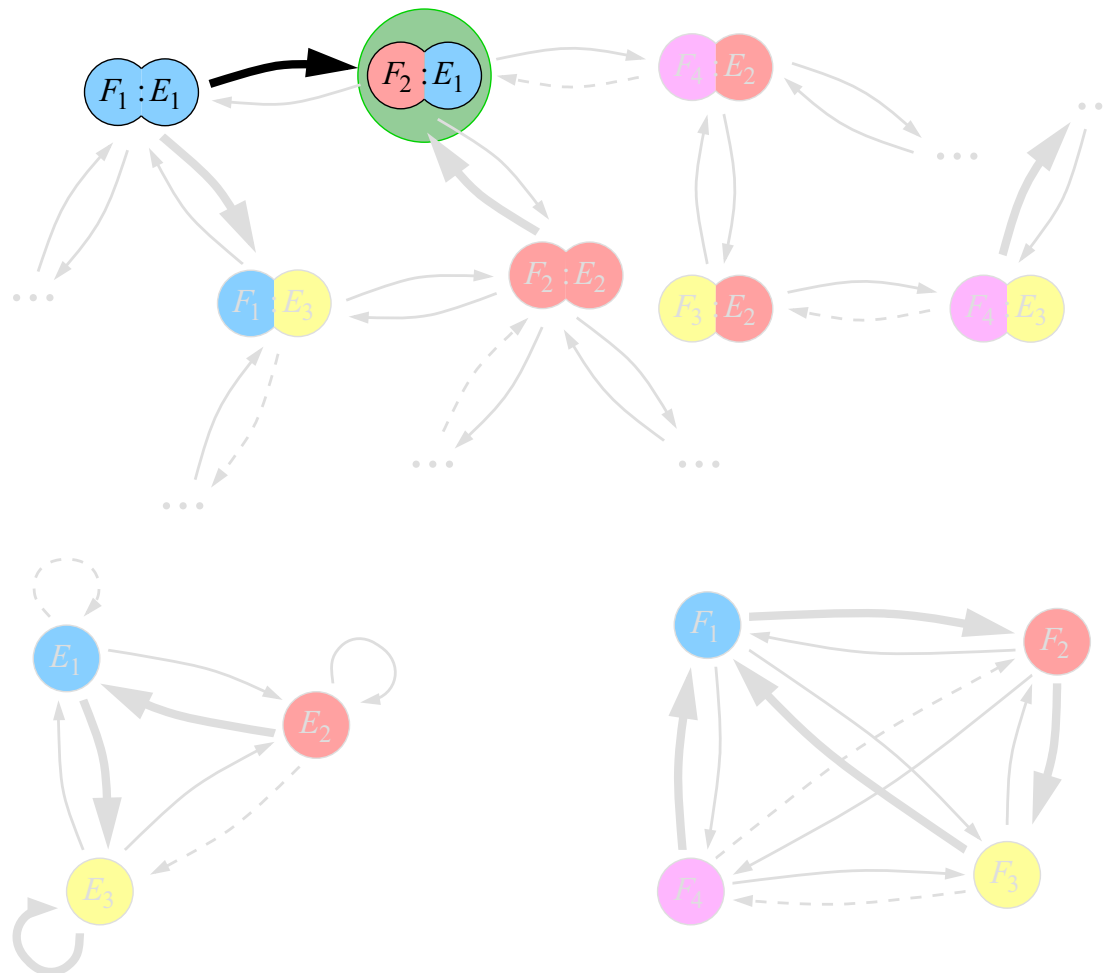
B: "Um 17 Uhr."

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte." 🤖

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin."

S: "Es tut mir leid. Wann möchten Sie abreisen?"

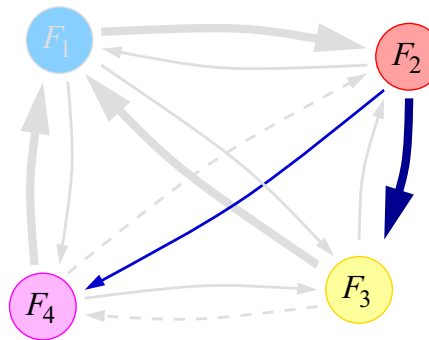
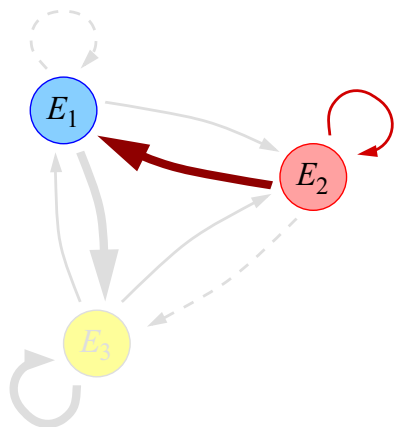
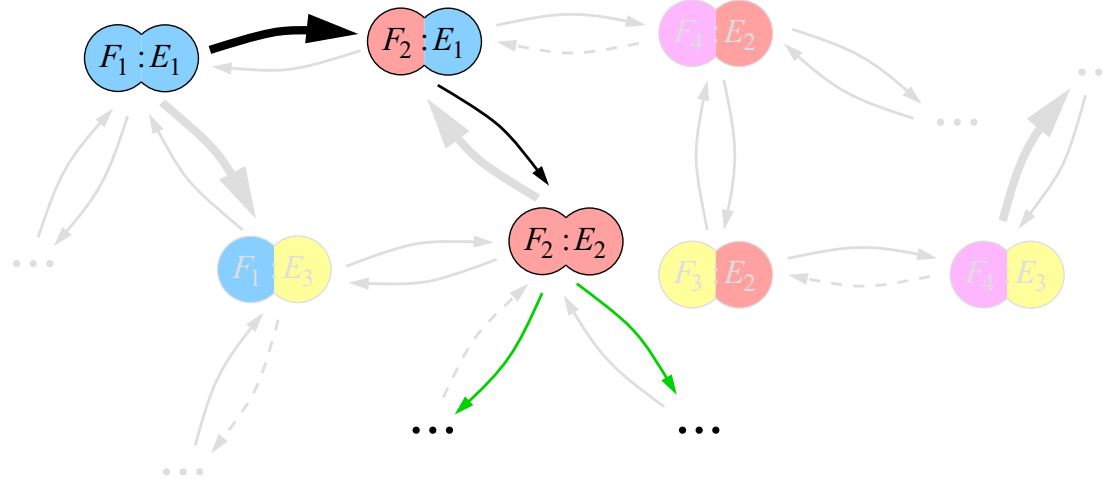
B: "Um 17 Uhr."

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte." 🤖

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin." 🤖

S: "Es tut mir leid. Wann möchten Sie abreisen?"

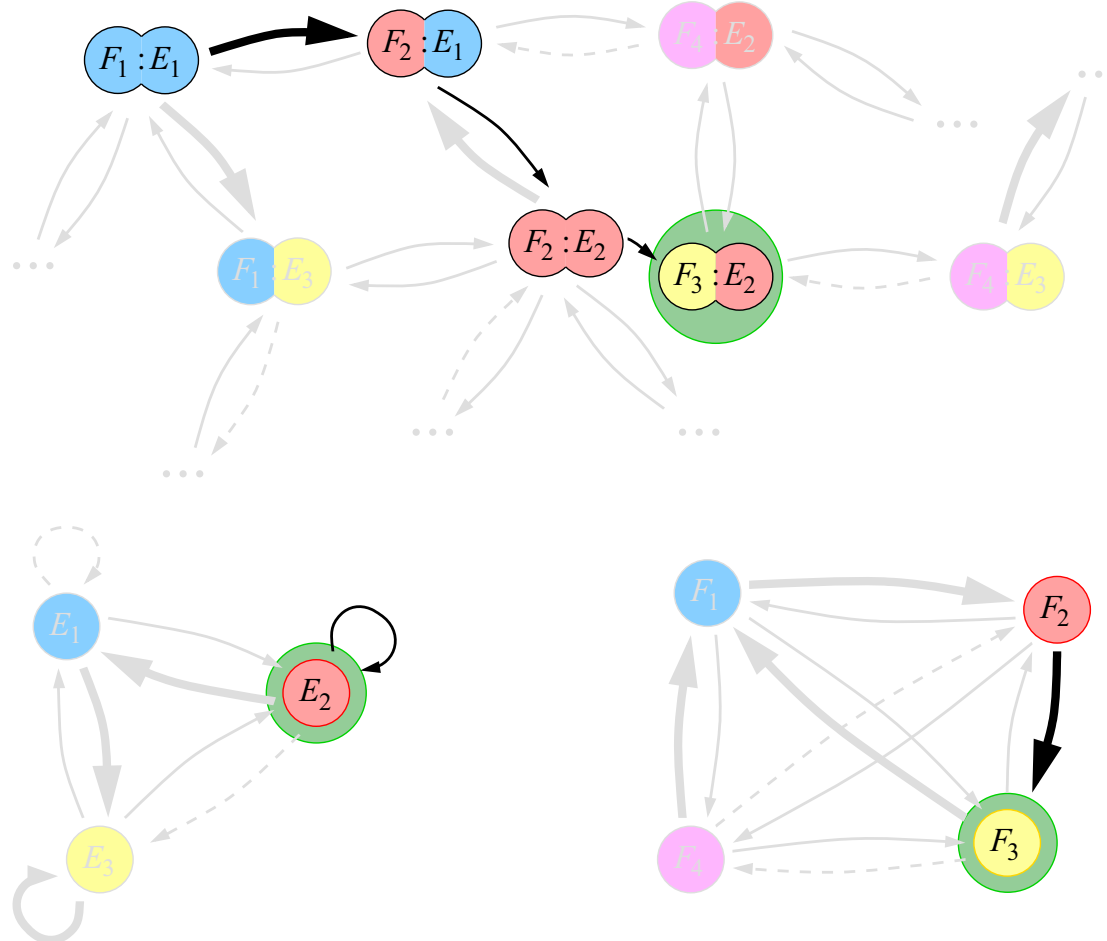
B: "Um 17 Uhr."

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte." 🤖

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin." 🤖

S: "Es tut mir leid. Wann möchten Sie abreisen?"

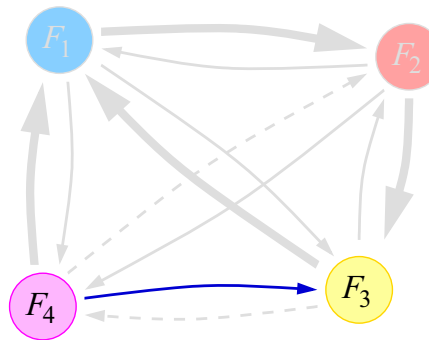
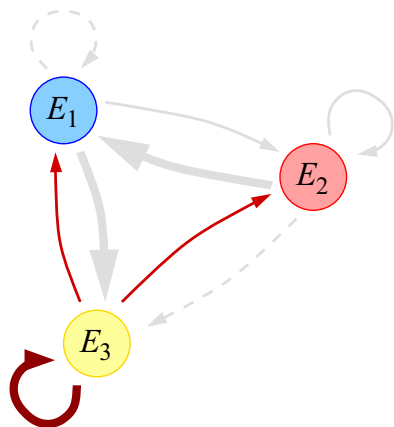
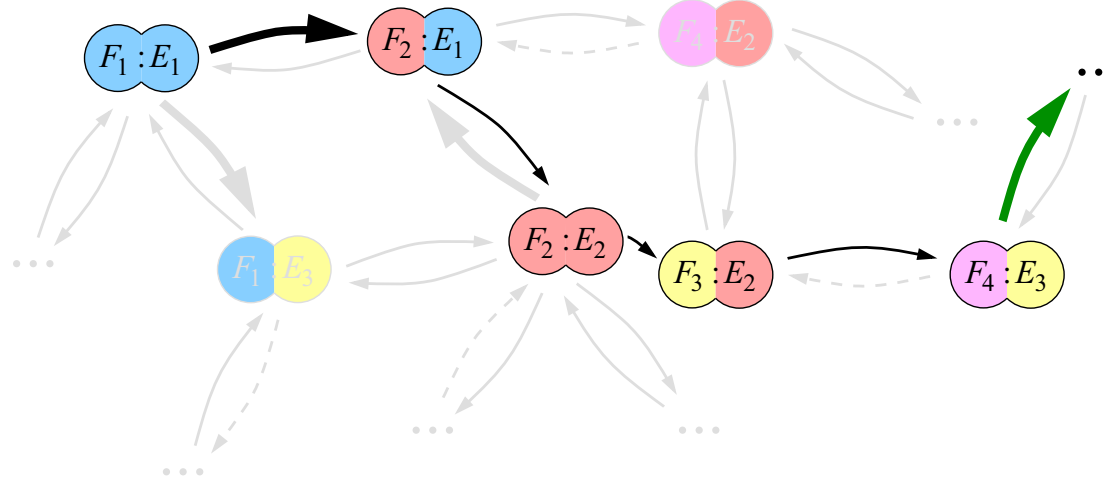
B: "Um 17 Uhr."

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte." 🙄

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin." 😞

S: "Es tut mir leid. Wann möchten Sie abreisen?"

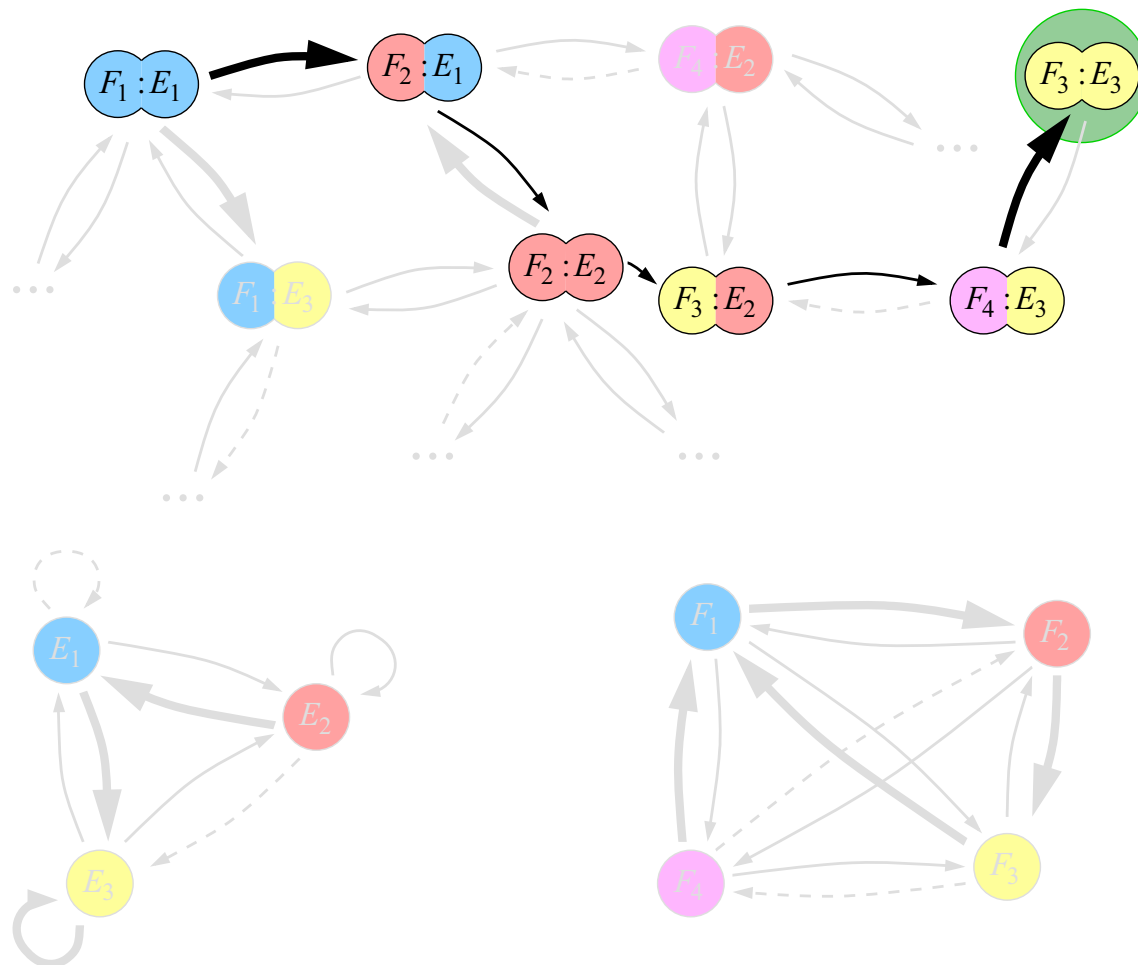
B: "Um 17 Uhr." 😊

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte." 🙄

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin." 😞

S: "Es tut mir leid. Wann möchten Sie abreisen?"

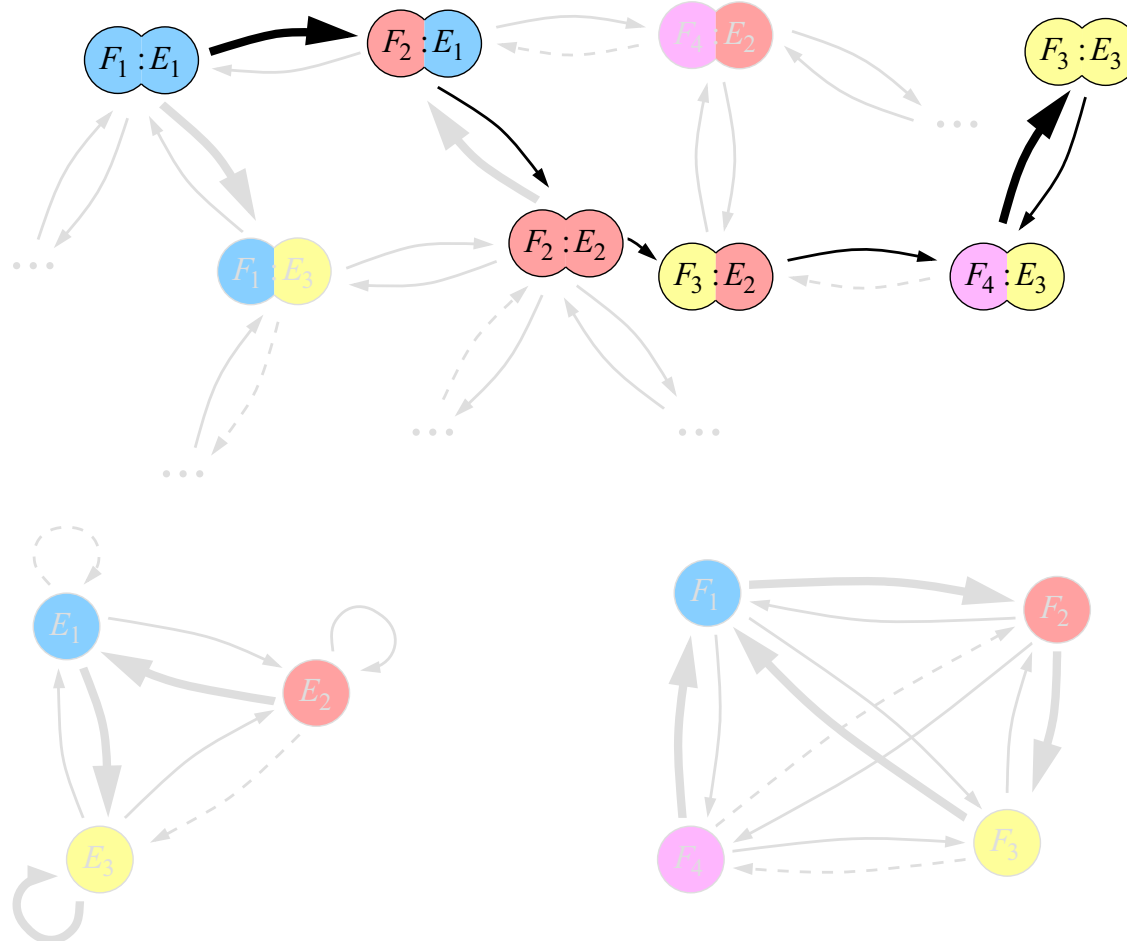
B: "Um 17 Uhr." 😊

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag"

S: "..."

Emotionales Dialogmodell



S: "... Wie kann ich Ihnen helfen?"

B: "Nach München bitte." 🙄

S: "Oh, von wo aus möchten Sie denn reisen?"

B: "Von Berlin." 😞

S: "Es tut mir leid. Wann möchten Sie abreisen?"

B: "Um 17 Uhr." 😊

S: "Schön, an welchem Tag möchten Sie reisen?"

B: "Am Freitag" 😊

S: "..."

Zusammenfassung: Wissenschaftlicher Beitrag

▶ Theoretische Aspekte

▶ Sprach-Emotionserkennung

▷ Synergien durch Kooperation

▷ Erkennerrleistung ↗ – Aufwand ↘

▶ ROVER-Algorithmus für mehrere Sprach-Emotionserkennung

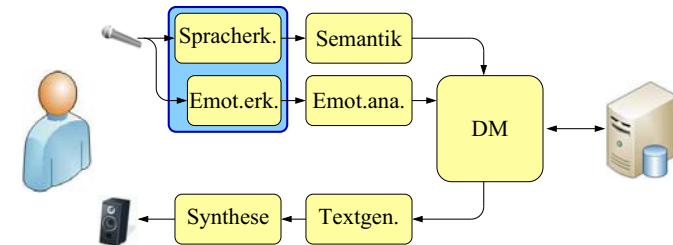
▶ Emotionales Dialogmodell – Parameterintegration

▶ Anwendungsbezogen & Experimentell

▶ Geeignete Features für Sprach-Emotionserkennung

▶ Emotionserkennerraten bis zu 76.4% bei 7 Emotionen

▶ Semantische Analyse zur Emotionserkennung



Zusammenfassung: Wissenschaftlicher Beitrag

▶ Theoretische Aspekte

▶ Sprach-Emotionserkennung

▷ Synergien durch Kooperation

▷ Erkennerleistung ↗ – Aufwand ↘

▶ ROVER-Algorithmus für mehrere Sprach-Emotionserkennung

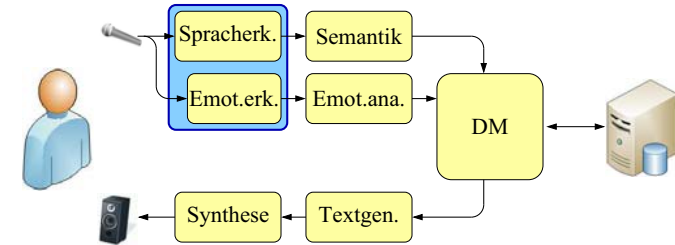
▶ Emotionales Dialogmodell – Parameterintegration

▶ Anwendungsbezogen & Experimentell

▶ Geeignete Features für Sprach-Emotionserkennung

▶ Emotionserkennungsraten bis zu 76.4% bei 7 Emotionen

▶ Semantische Analyse zur Emotionserkennung



Zusammenfassung: Wissenschaftlicher Beitrag

▶ Theoretische Aspekte

▶ Sprach-Emotionserkennung

▷ Synergien durch Kooperation

▷ Erkennerleistung ↗ – Aufwand ↘

▶ ROVER-Algorithmus für mehrere Sprach-Emotionserkennung

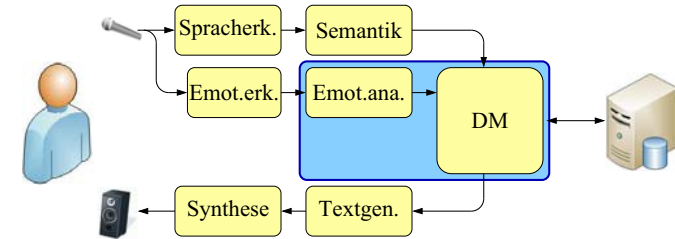
▶ Emotionales Dialogmodell – Parameterintegration

▶ Anwendungsbezogen & Experimentell

▶ Geeignete Features für Sprach-Emotionserkennung

▶ Emotionserkennungsraten bis zu 76.4% bei 7 Emotionen

▶ Semantische Analyse zur Emotionserkennung



Zusammenfassung: Wissenschaftlicher Beitrag

▶ Theoretische Aspekte

▶ Sprach-Emotionserkennung

▷ Synergien durch Kooperation

▷ Erkennerleistung ↗ – Aufwand ↘

▶ ROVER-Algorithmus für mehrere Sprach-Emotionserkenner

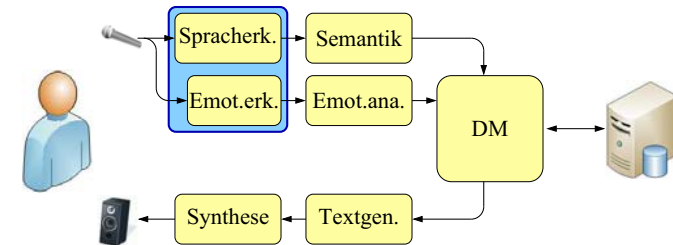
▶ Emotionales Dialogmodell – Parameterintegration

▶ Anwendungsbezogen & Experimentell

▶ Geeignete Features für Sprach-Emotionserkennung

▶ Emotionserkennerraten bis zu 76.4% bei 7 Emotionen

▶ Semantische Analyse zur Emotionserkennung



Zusammenfassung: Wissenschaftlicher Beitrag

▶ Theoretische Aspekte

▶ Sprach-Emotionserkennung

▷ Synergien durch Kooperation

▷ Erkennerrleistung ↗ – Aufwand ↘

▶ ROVER-Algorithmus für mehrere Sprach-Emotionserkennung

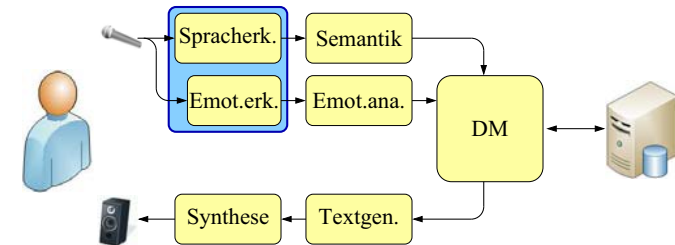
▶ Emotionales Dialogmodell – Parameterintegration

▶ Anwendungsbezogen & Experimentell

▶ Geeignete Features für Sprach-Emotionserkennung

▶ Emotionserkennerraten bis zu 76.4% bei 7 Emotionen

▶ Semantische Analyse zur Emotionserkennung



Zusammenfassung: Wissenschaftlicher Beitrag

▶ Theoretische Aspekte

▶ Sprach-Emotionserkennung

▷ Synergien durch Kooperation

▷ Erkennerleistung ↗ – Aufwand ↘

▶ ROVER-Algorithmus für mehrere Sprach-Emotionserkenner

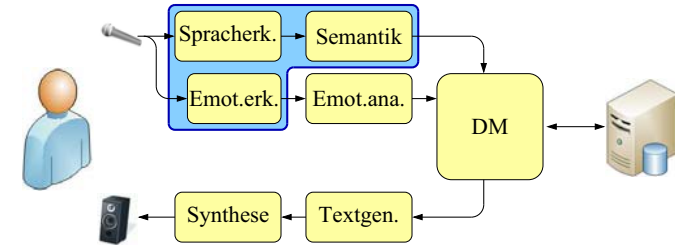
▶ Emotionales Dialogmodell – Parameterintegration

▶ Anwendungsbezogen & Experimentell

▶ Geeignete Features für Sprach-Emotionserkennung

▶ Emotionserkennerraten bis zu 76.4% bei 7 Emotionen

▶ Semantische Analyse zur Emotionserkennung

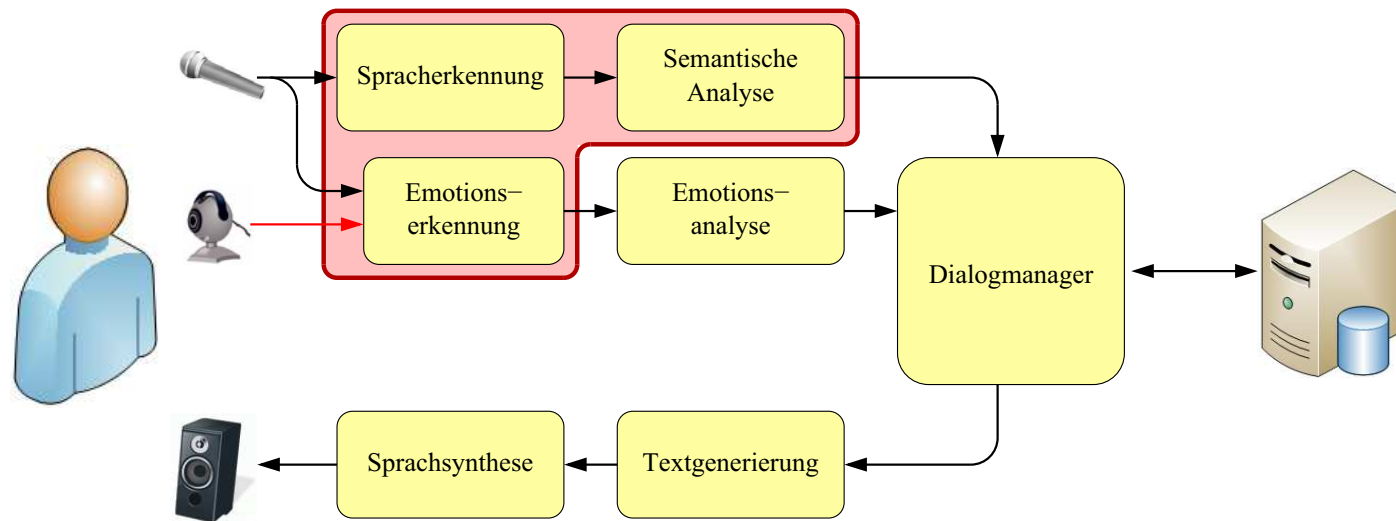


Ausblick

- ▶ Optimierung \Leftrightarrow wenig Trainingsdaten
 - ▷ Robustheit durch Backoff-Strategien
- ▶ Erweiterung des Systems
 - ▷ Emotionale Sprachsynthese
 - ▷ Weitere Modalitäten zur Emotionserkennung
- ▶ Evaluierung Dialogmodell & Gesamtsystem

Ausblick

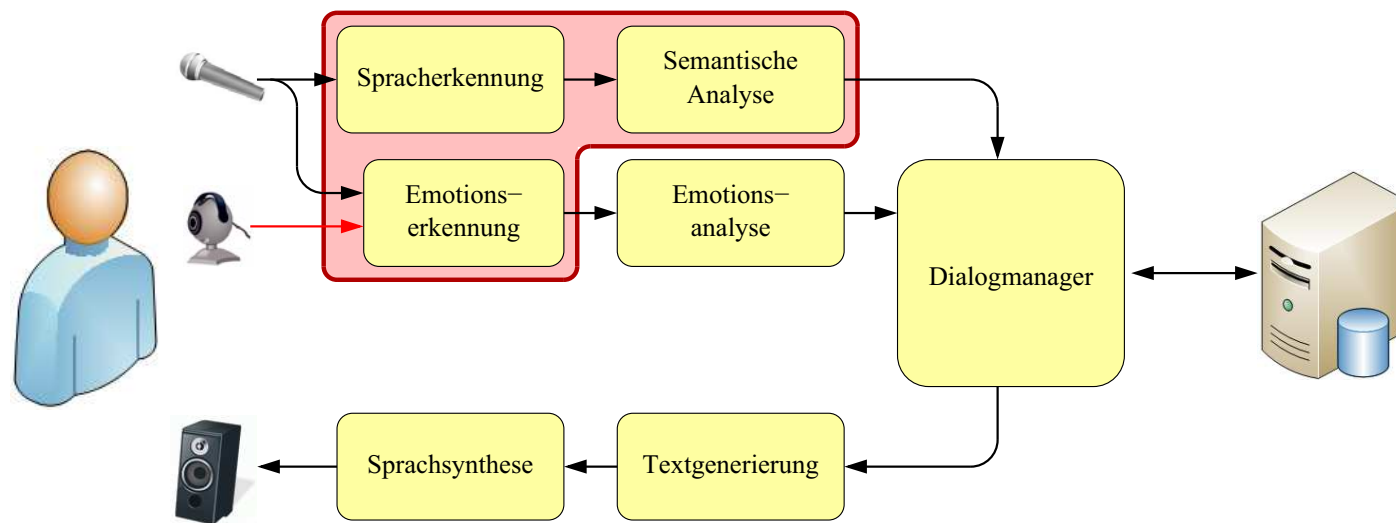
- ▶ Optimierung \Leftrightarrow wenig Trainingsdaten
 - ▷ Robustheit durch Backoff-Strategien
- ▶ Erweiterung des Systems
 - ▷ Emotionale Sprachsynthese
 - ▷ Weitere Modalitäten zur Emotionserkennung



- ▶ Evaluierung Dialogmodell & Gesamtsystem

Ausblick

- ▶ Optimierung \Leftrightarrow wenig Trainingsdaten
 - ▷ Robustheit durch Backoff-Strategien
- ▶ Erweiterung des Systems
 - ▷ Emotionale Sprachsynthese
 - ▷ Weitere Modalitäten zur Emotionserkennung



- ▶ Evaluierung Dialogmodell & Gesamtsystem

Veröffentlichungen

- ▷ Pittermann, J., Lentmaier, M., and Zigangirov, K. S. (2003). On Bandwidth-Efficient Convolutional LDPC Codes. In *IEEE International Symposium on Information Theory (ISIT)*, Yokohama, Japan.
- ▷ Pittermann, J., Lentmaier, M., and Zigangirov, K. S. (2005). On Coded Modulation Using LDPC Convolutional Codes. *AEÜ – International Journal of Electronics and Communications*, 59(4):254–257.
- ▷ Pittermann, J., Rittinger, A., and Minker, W. (2005). Flexible Dialogue Management in Intelligent Human-Machine Interfaces. In *The IEE International Workshop on Intelligent Environments*, Colchester, UK.
- ▷ Pittermann, J. (2005). Spoken Dialogue Technology: Toward the Conversational User Interface by Michael F. McTear. *Computational Linguistics*, 31(3):403–405.
- ▷ Pittermann, J. and Pittermann, A. (2006). Integrating Emotion Recognition into an Adaptive Spoken Language Dialogue System. In *2nd IET International Conference on Intelligent Environments*, volume 1, pages 197–202, Athens, Greece.
- ▷ Minker, W., Pittermann, J., Pittermann, A., Strauss, P.-M., and Bühler, D. (2006). Next-Generation Human-Computer Interfaces - Towards Intelligent, Adaptive and Proactive Spoken Language Dialogue Systems. In *2nd IET International Conference on Intelligent Environments*, Athens, Greece.
- ▷ Minker, W., Albalade, A., Bühler, D., Pittermann, A., Pittermann, J., Strauss, P.-M., and Zaykovskiy, D. (2006). Recent Trends in Spoken Language Dialogue Systems. In *ITI 4th International Conference on Information and Communications Technology (ICICT 2006)*, Cairo, Egypt.
- ▷ Pittermann, A. and Pittermann, J. (2006). Getting Bored with HTK? Using HMMs for Emotion Recognition. In *8th International Conference on Signal Processing (ICSP)*, pages 704–707, Guilin, China.
- ▷ Pittermann, J. and Pittermann, A. (2006). A Post-Processing Approach to Improve Emotion Recognition Rates. In *8th International Conference on Signal Processing (ICSP)*, pages 708–711, Guilin, China.
- ▷ Pittermann, J. and Pittermann, A. (2006). An 'Emo-Statistical' Model for Flexible Dialogue Management. In *8th International Conference on Signal Processing (ICSP)*, pages 1599–1602, Guilin, China.
- ▷ Pittermann, J. and Pittermann, A. (2007). A Data-Oriented Approach to Integrate Emotions in Adaptive Dialogue Management. In *International Conference on Intelligent User Interfaces (IUI)*, pages 270–273, Honolulu, USA.

Veröffentlichungen

- ▷ Abdennaher, S., Aly, M., Bühler, D., Minker, W., and Pittermann, J. (2007). BECAM Tool – A Semi-automatic Tool for Bootstrapping Emotion Corpus Annotation and Management. In *European Conference on Speech and Language Processing (EUROSPEECH)*, Antwerp, Belgium.
- ▷ Pittermann, J., Pittermann, A., Meng, H., and Minker, W. (2007). Towards an Emotion-Sensitive Spoken Dialogue System – Classification and Dialogue Modeling. In *3rd IET International Conference on Intelligent Environments*, Ulm, Germany.
- ▷ Aly, M., Pittermann, J., Bühler, D., and Minker, W. (2007). A Semi-Automatic Tool for Emotion Corpora Annotation. In *3rd IET International Conference on Intelligent Environments*, Ulm, Germany.
- ▷ Pittermann, J., Minker, W., Pittermann, A., and Bühler, D. (2007). ProblEmo – Problem Solving and Emotion Awareness in Spoken Dialogue Systems. In *3rd IET International Conference on Intelligent Environments*, Ulm, Germany.
- ▷ Meng, H., Pittermann, J., Pittermann, A., and Minker, W. (2007). Combined Speech-Emotion Recognition for Spoken Human-Computer Interfaces. In *IEEE International Conference on Signal Processing and Communications (ICSPC)*, Dubai, UAE.
- ▷ Pittermann, J., Pittermann, A., and Minker, W. (2007). Design and Implementation of Adaptive Dialogue Strategies for Speech-Based Interfaces. *Journal of Ubiquitous Computing and Intelligence*, 1(2):145–152.
- ▷ Pittermann, J., Schmitt, A., and Fawzy El Sayed, N. (2008). Integrating Linguistic Cues into Speech-Based Emotion Recognition. In *4th IET International Conference on Intelligent Environments*, Seattle, USA.
- ▷ Pittermann, J., Pittermann, A., Schmitt, A., and Minker, W. (2008). Comparing Evaluation Criteria for (Automatic) Emotion Recognition. In *4th IET International Conference on Intelligent Environments*, Seattle, USA.
- ▷ Minker, W., Pittermann, J., Pittermann, A., Strauss, P.-M., and Bühler, D. (to appear 2008). Challenges in Speech-Based Human-Computer Interfaces. *International Journal of Speech Technology (IJST)*.
- ▷ Minker, W., Pittermann, J., Pittermann, A., Strauss, P.-M., and Bühler, D. (to appear 2008). Intelligent and Empathic Speech Interfaces. In Yablonski, R. H., editor, *Speech Communication at the Leading Edge*, pages 73–107. Nova Science Publishers, Inc., Hauppauge, USA.

Herzlichen Dank für Ihre Aufmerksamkeit!