

Inbetriebnahme, Konfigurierung und Evaluation des Vorverarbeitungsmoduls zur Segmentierung von Dialogakten mittels Pausenerkennung

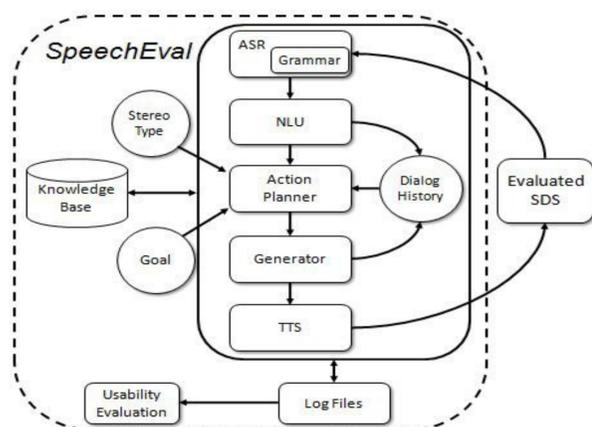
Bachelorarbeit Katarina Birghan

Aufgabe:

Die Zielstellung dieser Bachelorarbeit war die Implementierung eines Moduls zur automatischen Dialogaktsegmentierung im Rahmen der Evaluierung von Sprachdialogsystemen.

Gegenwärtig wird die Evaluierung mit einer Anzahl von Testpersonen durchgeführt. Dabei werden Kriterien wie Effizienz, Komfort, Gebrauchstauglichkeit und Akzeptanz untersucht. Dies ist recht kosten- und zeitintensiv.

Zur automatischen Evaluierung wird im DFKI-Forschungsprojekt SpeechEval eine Simulationsumgebung entwickelt. Das Segmentierungsmodul ist dabei Bestandteil der NLU-Komponente.



Architektur des SpeechEval Systems

Die Simulationsumgebung wird über eine Menge von Wissen über verschiedene Benutzertypen verfügen. Aus einem Korpus von Testdialogen mit verschiedenen kommerziellen Sprachdialogsystemen wird deshalb typisches Benutzerverhalten abgeleitet. Dafür müssen die Äußerungen (Sätze) der Dialogpartner in einzelne Dialogakte segmentiert und entsprechend klassifiziert werden.

Der Dialogakt ist die Basiseinheit eines Dialoges. Er beschreibt die Handlung, die von einem Sprecher mit einer Äußerung ausgeht und hat eine Funktion bei der Dialogzielerreichung. Das kann beispielsweise eine Frage oder eine Begrüßung sein.

Die Dialogaktklassifizierung ist abhängig vom Kontext.

Turn	Äußerung	Dialogakt	Funktion
No Speaker	#NS	NOICE	nichtsprachliche Laute/Hintergrundgeräusche
SDS:	Herzlich Willkommen bei OrangeWings.	HELLO	Begrüßung
	Sie sprechen mit Resa, dem automatischen Reservierungssystem von OrangeWings. Sie können bei mir ein Zimmer mit Kreditkarte reservieren oder mit dem nächsten freien Mitarbeiter sprechen. Sagen Sie bitte Resa oder Mitarbeiter.	PROVIDE_INFO	Das System teilt Informationen mit
		ALTERNATIVE_QUESTION	Anrufer soll eine Option auswählen.

Beispiel einer Dialogaktsegmentierung

Zur Segmentierung wurde das prosodische Merkmal Pause genutzt.

In der menschlichen spontanen Sprache gehen aufeinanderfolgende Wörter ineinander über. Am Ende von Sätzen sind Pausen länger und treten häufiger auf als innerhalb von Sätzen. Oft sind auch erst nach mehreren Sätzen Pausen deutlich wahrnehmbar.

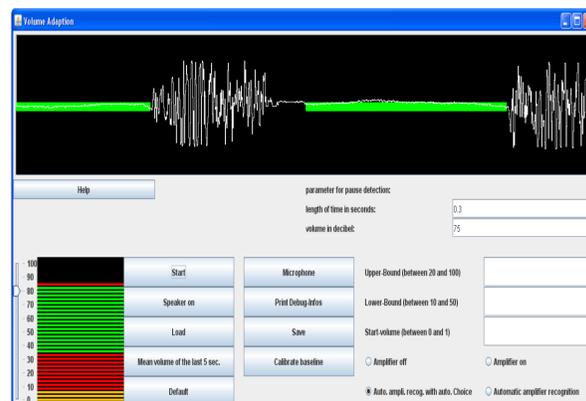
Aus linguistischer Sicht haben Pausen strukturelle Funktionen und lösen Doppeldeutigkeiten auf.

Deshalb eignet sich die Pause für die Segmentierung von Dialogakten.

Pausen lassen sich direkt durch Untersuchung des Verlaufs der Lautstärke des Sprachsignals detektieren.

Implementierung:

Zur Lösung der Aufgabe wurde ein vorhandenes Vorverarbeitungsmodul erweitert. Dieses Modul erkannte Sprechpausen in einem eingehenden Sprachsignal und zeigte diese grafisch an. Außerdem verfügte es über die Funktion der Lautstärkennormalisierung.



GUI des Vorverarbeitungsmoduls

Die Erkennung der Pausen wird durch die Parameter Lautstärke in dB und Dauer in sek gesteuert. Der Parameter Lautstärke stellt dabei einen Schwellwert dar, anhand dessen entschieden wird, ob ein Sprachabschnitt als Pause oder als Sprache klassifiziert wird. Mittels des Parameters Dauer wird nur dann eine Pause markiert, wenn sie eine bestimmte Mindestlänge hat. Die erkannten Pausen können zwischen einzelnen Wörtern, Sätzen oder mehreren Sätzen liegen.

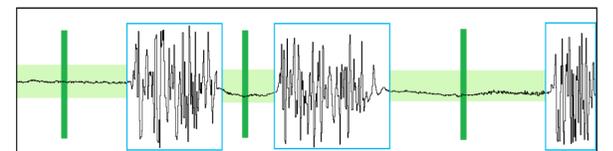
In dieser Arbeit bestand die Aufgabe, herauszufinden, welche Werte die beiden Parameter annehmen müssen, damit nur die Pausen, die zwischen Dialogakten liegen, erkannt werden.

Dazu wurden verschiedene Parametereinstellungen untersucht und auf ihre Eignung zur Segmentierung in einer Evaluierungsphase bewertet.

Das Vorverarbeitungsmodul wurde um die Funktion erweitert, eine Menge von Dialogaufzeichnungen des VOICE Awards Korpus hintereinander abzuspielen und jede einzelne Datei unter einer Menge von Parameterkombinationen auf Pausen zu untersuchen.

Zur Bewertung der Effektivität einer Parameterkombination wurden die Standard-Evaluierungsmetriken Recall, Precision und F-Measure verwendet.

Die Evaluierung erfolgte durch Vergleich der vom Modul erkannten Pausen mit der korrekten Segmentierung aus manuellen Annotationen.



Beispiel einer Übereinstimmung der Positionen von erkannten und tatsächlichen Pausen zwischen den Dialogakten

Dazu wurde ein Evaluierungsmodul in Java implementiert.

Evaluierung:

In einer ersten Testphase wurde untersucht, ob sich ungünstige Parameterkombinationen ausschließen lassen. Dazu wurden 10 Dialoge aus 10 verschiedenen Sprachdialogsystemen auf Pausen untersucht und die Ergebnisse bewertet.

In der Testphase konnte ein Ranking der untersuchten Kombinationen festgestellt werden.

Parameterkombination	Recall	Precision	F-Measure
0.8 - 85	42%	53%	46%
0.8 - 80	42%	65%	51%
0.8 - 75	41%	57%	50%
0.7 - 85	39%	52%	42%
0.7 - 80	40%	54%	47%
0.7 - 75	34%	55%	40%
0.6 - 85	34%	51%	41%
0.6 - 80	36%	53%	41%
0.6 - 75	32%	47%	39%

Ranking der Parameter

Es zeigte sich, dass sich die Parametereinstellung mit einer Lautstärke von 80 dB und einer Dauer von 0.8 Sekunden am besten für den VOICE Awards Korpus eignet, da bei dieser Einstellung der beste F-Measure erreicht wird.

Diese Parametereinstellung wurde anschließend für die eigentliche Evaluierungsphase verwendet. Hierfür standen 661 Dialoge von verschiedenen Sprachdialogsystemen zur Verfügung.

Als Ergebnis wurde ein durchschnittlicher F-Measure von 52% erreicht.

Im Einzelnen unterschieden sich die Ergebnisse von Dialogsystem zu Dialogsystem. Einige Systeme haben einen F-Measure von 92% und andere wiederum liegen bei 0%.

Fazit:

Bedingt durch die unterschiedliche Audioqualität der einzelnen Dateien, wie große Lautstärkenunterschiede, starkes Rauschen und Hintergrundgeräusche wurden unterschiedliche Ergebnisse erreicht.

Anzunehmen ist, dass mit einer optimierten Lautstärkennormalisierung bessere Ergebnisse erreicht werden können.

Eine weitere Verbesserung kann durch Implementierung einer Rauschentfernung und durch die Einbeziehung des Rauschgrenzwertes des Rechners erreicht werden. Zusätzlich können weitere prosodische Merkmale wie Grundfrequenz und F0 genutzt werden.