

Spracherkennung mit 3D- Videostream

Abschlusspräsentation

Gruppe:

S. Leon
A. Klein

Themenvorstellung

- Aufgabe
 - Erkennung einer Sprechenden Person mit Hilfe eines RGBD Sensor
- Verwendete Hardware und Software
 - RGBD Sensor Asus Xiton Pro
 - Faceshift
 - Visual Studio
- Herausforderungen
 - Auswertung des Netzwerkstream von Faceshift
 - Analyse von Daten

Asus Xiton Pro

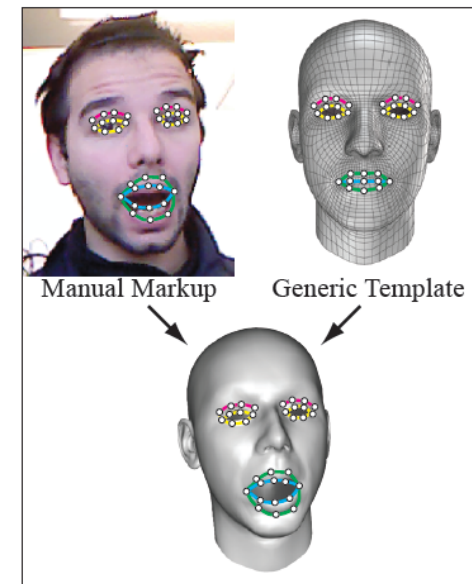


Technische Daten:

- Auflösung: 1280x1024
- Tiefenauflösung: 640x480
- Distanz: 0.8m bis 3.5m

Faceshift

- Software zur Animierung von 3d Gesichtern
- Funktionsweise:
 - RBD-Videostream
 - Algorithmus basierend auf Video und Tiefen-Daten
 - Fitting Prozess mit Hilfe von Blendshapes/Action Units
- Ermöglicht die Erfassung von:
 - Gesichtszügen
 - Kopfdrehung
 - Blickwinkel



Gedanken zum Programm

- Möglichkeiten von Faceshift zeigen
- Wieso Spracherkennung mit RGBD-Sensor
 - Wer spricht Gerade(Gesichtserkennung)?
 - Mehrere Quellen detektieren
- Einfacher als Emotionen erkennen(Weniger Daten)
- Forschungsthema (Actionunits)

Programmfunktionen

- Spracherkennung
 - Mit vorausgehendem Einlernverfahren
 - Einlernverfahren während dem Betrieb
- Kopf und Gaze Ausrichtung
- Aufnahme der 51 Blendshapes

Funktionsweise

- 10 Blendshapes aus Bereich Mund und Unterkeifer
- Berechnung der Koeffizienten
 - Über gleitenden Mittelwert der Änderung

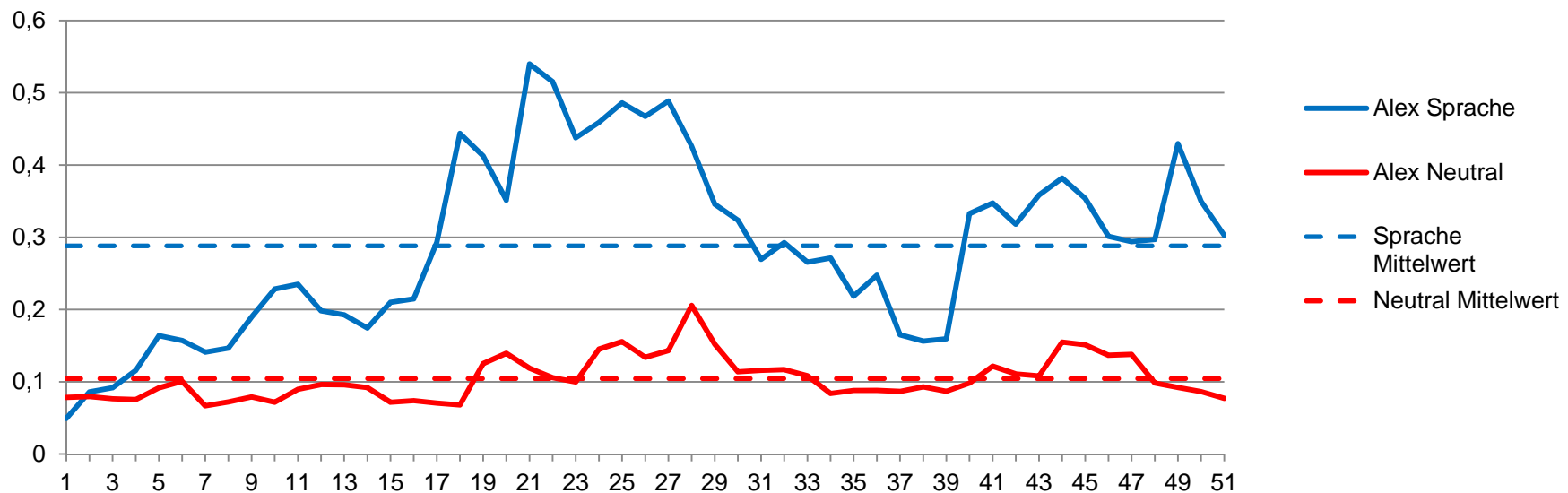
f(k) = Gleitender Mittelwert

g(k) = Coeffizientenwerte

$$f(k) = \sum_{i=0}^{10} |\dot{g}(k - 1)|$$

Funktionsweise

- 2 Verfahren zur Auswertung
 - Während des Betriebs
 - Einlernverfahren: Vergleich Sprache zu Neutral
- Faktoren aus Analyse mit Excel



Programm Demo

Zukünftige Aufgaben

- Sprachlernverfahren
- Andere Verfahren als Mittelwertbildung
- Prüfung der Faktoren
- Datenbank erstellen
- Pose Normalization Automatisch

Fazit

Vielen Dank für Ihre
Aufmerksamkeit

Noch Fragen?