

Jarek Krajewski¹, Rainer Wieland¹
Bergische Universität Wuppertal / Wirtschaftspsychologie

Stimmakustische Detektion von arbeitssicherheitskritischen Befindlichkeitszuständen

1 Einsatzfelder stimmakustischer Warnsysteme

20% aller tödlichen Verkehrsunfälle werden auf Schläfrigkeit zurückgeführt (Garbarino et al., 2001). Darüber hinaus bilden sicherheitsrelevante Befindlichkeitszustände wie z.B. Angst, Stress, Wut, Schmerz, Monotonie oder Alkoholisiertheit ein zusätzliches Unfallrisiko. Präventionsansätze setzen daher auf die Identifikation von Personen, die diese sicherheitskritischen Zustände mit erhöhter Wahrscheinlichkeit zeigen (siehe z.B. MPU). Trotz entsprechender Eignungsuntersuchungen kommt es im Straßenverkehr jedoch zu einer immensen Zahl an Unfallereignissen. Daher stellt die situative Detektion aktueller kritischer Zustände über die Entwicklung von *Schnelltests* oder *automatischer Echtzeit-Zustands-Warnsysteme* (Hagenmeyer, 2007) aus verkehrs- und arbeitswissenschaftlicher Präventionsperspektive eine wichtige Optimierungsmöglichkeit und Herausforderung dar (vgl. Abb. 1).

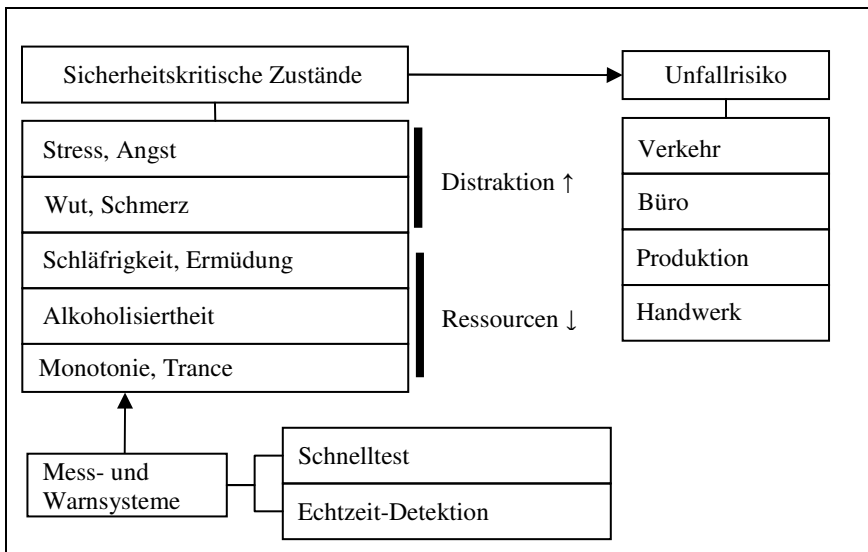


Abb. 1: Rahmenstruktur zur Identifikation sicherheitskritischer Zustände

Der Analyse sprachlich-akustischer Informationen zur Bestimmung des Befindlichkeitszustandes ist schon seit den 60er Jahren anvisiert. Aber erst die Fortschritte der Prozessortechnik, Mustererkennung und sowie der Transfer von Kennzahlen der Sprach- bzw. Sprechererkennung ermöglichen die Analyse von satzlängigen Sprachaufnahmen. Im Gegensatz zu physiologischen Messansätzen bietet die akustische Stimmanalyse, die Vorzüge eines *berührungs- und kalibrierungsfreien, ökonomischen sowie verfälschungsresistenten Messzugangs*. Der laufende Tätigkeitsvollzug wird nicht beeinträchtigt, da eine „Hands-free“ und „Eyes-free“ Messsituation vorliegt. Stimmakustische Warnsysteme können als Schnelltest („Fit-for-Duty-Tests“) oder auch als Online-Detektionssystem konzipiert sein. In kommunikationsintensiven Berufen (z.B. Fluglotsen, Call Center Agent, Lehrer) ist die stimmakustische Online-Analyse aufgrund des allseits verfügbaren Sprachmaterials denkbar. In anderen - auch mobilen - Messsituationen (wie z.B. im Straßen-Schiene-Luftverkehr, Handwerk, Büro oder der Produktion) sind Schnelltests mit 5 Sek. langen Testsätzen vorstellbar. Die Anforderungen an diese Schnelltests sind im Falle von Schläfrigkeitsdetektionsmodulen vor allem eine niedrige Falsch-Alarmrate ($< 1\%$) und eine hohe Benutzerakzeptanz ($> 70\%$).

2 Akustisches Messprozedere

Um die mit den sicherheitskritischen Zuständen assoziierten Stimmveränderungen abzubilden, werden breite akustische Mehrzweck-Kennzahlensets bestimmt. Diese setzen sich aus Kennzahlen zusammen, die *Intensitäts-, Intonations-, Rhythmus-, Pausen-, und Stimmqualitätsphänomene* abbilden. Darüber hinaus wird das Kennzahlen-Set ergänzt durch Resonanzfrequenzmaßzahlen (Formanten), die Auskunft über Artikulationsphänomene und weitere physiologische Veränderungen des Vokaltrakts liefern. So repräsentieren Formantposition und -bandbreite neben der Position und Bewegung von Artikulatoren (z.B. Kieferwinkel, horizontale Zungenposition, Lippenstülpung, Verengung des Vokaltrakts und Velumsenkung) auch Vokaltrakteigenschaften wie Krümmungsgrad, Wärmeabstrahlung, Reibung oder Schwingungseigenschaften der Vokaltraktwände). Des Weiteren sind nicht nur Vokaltrakteigenschaften sondern auch Anspannungszustände des Brustbereichs mit den oben genannten Energieverlusten und daher Formantveränderungen assoziiert. Einen Überblick zu gängigen Kennzahlenfamilien wie sie in der Speech Emotion Recognition eingesetzt werden bietet Batliner et al. (2006) oder Schuller et al. (2007).

Die dort genannten akustischen Kennzahlen werden über einen in der künstlichen Intelligenzforschung üblichen und sich zunehmend auch in der Biosignalanalyse etablierenden Mustererkennungsprozess (z.B. Golz et al., 2007) zu einem Messwert von Schläfrigkeit, Wut, Angst, Alkoholisiertheit oder Stress zusammengefasst. Dieser Mustererkennungsprozess beinhaltet die Phasen: (a) Signalaufzeichnung (44,1 kHz; 16 Bit), (b) Vorverarbeitung (Artefaktkorrektur; Segmentation; Signal-Fensterung), (c) Merkmalsextraktion (58 akustische Kennzahlenverläufe; Veränderungs- und Beschleunigungskonturen; temporale Konturbeschreibungsgrößen, z-Normalisierungen = 40.000 *akustische Kennzahlen*), (d) Dimensionalitätsreduktion (Principle Component Analysis; Korrelationsfilter; Wrapper-basierte Merkmalsselektion), (e) Klassifikation (Neuronale Netze; Support Vector Machine; Metaklassifikation) und (f) Validierung (10-fache Kreuzvalidierung; Leave-one-sample-out) (vgl. Krajewski & Kröger, 2007; Krajewski, Wieland & Batliner, 2008).

Zahlreiche Studien belegen die Validität der akustischen Stimmanalyse in der Detektion von sicherheitskritischen interner Zuständen wie Stress, Wut, Angst, Alkoholisiertheit und Schläfrigkeit (z.B. Batliner et al., 2006; Vlasenko et al. 2007; Zhou, Hansen & Kaiser, 1999). Die Vorhersagegüten von Angst, Aggression und Schläfrigkeit liegen derzeit für ungesehene Sprecher bei ca. 80 % *Klassifikationsgenauigkeit* (z.B. Krajewski & Kröger, 2007; Schuller et al., 2007; Batliner & Huber, 2007). Illustrierend sind für den Fall der Schläfrigkeitsdetektion in Abb. 2 zwei häufig verwendete Kennzahlenverläufe dargestellt.

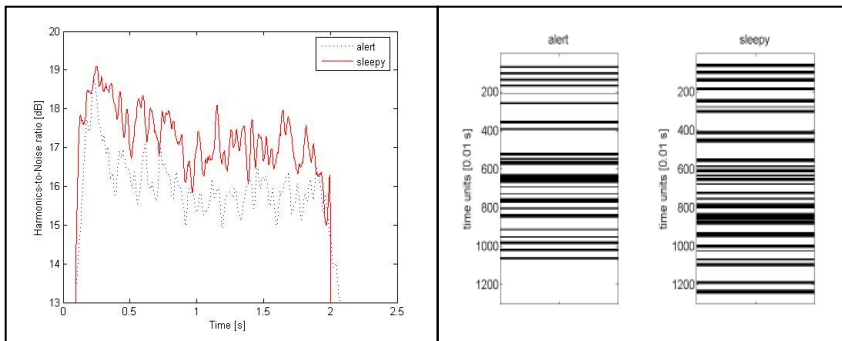


Abb. 2: HNR Konturen für wache vs. Mikroschlaf gefährdete Probanden (strichlierte vs. durchgezogene Linie) (links); Sprechpausenmuster (Pause= dunkler Balken) wacher vs. Mikroschlaf gefährdete Probanden (rechts)

3 Einschätzung der Anwendungsreife

Die Vorhersagegüten von sicherheitskritischen Zuständen wie Angst, Alkoholisiertheit, Aggression und Schläfrigkeit liegen derzeit für ungesehene Sprecher bei ca. 80 % Klassifikationsgenauigkeit. Verbessert werden kann die Messung um ca. 10% Klassifikationsgenauigkeit, wenn auf festgelegte Testsätze, relativ ruhige Umgebungsbedingungen und auf Trainings Sprachproben des Sprechers zurückgegriffen werden kann. Diese Rahmenbedingungen entsprechen den Anforderungen vieler Schnelltestsituationen, in denen ein dem System bekannter Sprecher einen 5-10 Sek. langen Testsatz spricht und ohne weiteren Aufwand innerhalb von einer Minute vollautomatisch ein Messergebnis zurückgemeldet bekommen muss. Damit die - sich derzeit im proof-of-concept Stadium befindenden - Detektions-Algorithmen diese Anforderungen an mobile, handliche und einfach zu bedienende Messinstrumenten erfüllen, müssen in Zukunft auf Mobiltelefon- bzw. PDA-Basis arbeitende Implementierungslösungen entwickelt werden.

Literatur

- Batliner, A. & Huber, R. (2007). Speaker Characteristics and Emotion Classification. In C. Müller (Ed.) *Speaker Classification I Fundamentals, Features, and Methods* (pp. 138-151). LNAI Berlin-Heidelberg: Springer.
- Garbarino, S., Nobili, L., Beelke, M., De Carli, F. & Ferrillo, F. (2001). The contributing role of sleepiness in highway vehicle accidents. *Sleep*, 24, 203-206.
- Golz, M., Sommer, D., Chen, M., Trutschel, U., Mandic, D. (2007). Feature Fusion for the Detection of Microsleep Events; *J VLSI Signal Proc Syst*, 49, 329-342.
- Hagenmeyer, L. (2007). *Development of a Multimodal, Universal Human-Machine-Interface for Hypovigilance-Management-Systems*. Heimsheim: Jost-Jetter-Verlag.
- Krajewski, J., Kröger, B. (2007). Using prosodic and spectral characteristics for sleepiness detection. *Interspeech Proceedings*, 1841-1844.
- Krajewski, J., Wieland, R. & Batliner, A. (2008). An acoustic framework for detecting fatigue in speech based Human-Computer-Interaction. *11th International Conference on Computers Helping People with Special Needs Linz, Austria, July 9-11*.
- Schuller, B., Batliner, A., Seppi, D., Steidl, S., Vogt, T., Wagner, J., Devillers, L., Vidrascu, L., Amir, N., Kessous, L., Aharonson, V. (2007). The Relevance of feature type for the automatic classification of emotional user states: Low level descriptors and functionals. *Proceedings of Interspeech*, 2253-2256.
- Vlasenko, B., Schuller, B., Wendemuth, A., Rigoll, G. (2007). Combining frame and turn-level information for robust recognition of emotions within speech. *Proceedings of Interspeech*, 2249-2252.
- Zhou, G., Hansen, J.H.L. & Kaiser, J.F. (1999). Methods for Stressed Speech Classification: Nonlinear TEO and Linear Speech Based Features. *IEEE International Conference on Acoustics, Speech, Signal Processing*, 4, 2087-2090.