

Schnarcherkennung mit diskreten Hidden-Markov-Modellen

Benjamin Kraus, Stefan Feldes

HS Mannheim Institut für Digitale Signalverarbeitung, 68163 Mannheim,
E-Mail: benjamin.kraus@gmx.com, s.feldes@hs-mannheim.de

Einleitung

Schnarchen ist ein weitverbreitetes Phänomen, das die Nachtruhe des Betroffenen und der Bettnachbarn beeinträchtigt. Die Folgen reichen von einem simplen Ärgernis bis hin zu erhöhten Risiken von Bluthochdruck und Gefäßerkrankungen. Behandlungsmethoden für schwerwiegende Fälle umfassen beispielsweise eine maschinelle Unterstützung der Atemluftaufnahme oder operative Eingriffe am weichen Gaumen und Zungenboden. Um den Erfolg solcher Behandlungen zu überprüfen oder nicht-invasive Methoden, wie etwa eine sofortige Beeinflussung des Schnarchenden mittels eines Vibrationsarmbandes [1], zu ermöglichen, ist es notwendig Schnarchen in Echtzeit mit einfachen Mitteln zu detektieren. Eine mögliche Vorgehensweise ist die Detektion des Schnarchgeräusches in Audioaufnahmen des Schlafenden. Hierzu wurden in [2] kontinuierliche Hidden-Markov-Modelle (HMM) vorgeschlagen und deren Potential aufgezeigt. Der vorliegende Beitrag beschreibt die Realisierung eines aufwandsarmen, auf diskreten HMM basierenden Lösungsansatzes und dessen Validierung durch Experimente.

Lösungsansatz

Der vorgeschlagene Ansatz basiert im Wesentlichen auf der Übertragung von aus der Spracherkennung bekannten Methoden [3]. So wurden zur Erfassung der spektralen Einhüllenden und Energie die vielfach bewährten Mel-Frequency-Cepstral-Coefficients (MFCC) gewählt. Ein Vektorquantisierer wird angewendet, um die Analyse durch diskrete HMM zu ermöglichen. Die Eigenschaften des Schnarchens werden in einem HMM statistisch modelliert, die Umgebungsgeräusche in einem zweiten HMM. Beide HMM werden miteinander verknüpft, um ein Modellsystem zu erzeugen, das die Analyse von fortlaufenden Audiosignalen, die Anteile beider Geräuschklassen beinhalten, ermöglicht. Die beiden HMM wurden mit dem im Folgenden erläuterten Trainingsprozess erstellt.

Training

Zum Training standen ein Datensatz mit 931 Schnarchproben aus der Universitäts-HNO-Klinik Mannheim sowie ein Datensatz mit 468 Schnarch- und 363 Nebengeräuschen aus häuslichem Umfeld zur Verfügung. In einem Vorverarbeitungsschritt wurden diese Proben durch Zero-padding in ihrer Dauer vereinheitlicht. Die spektralen Merkmale wurden in Form von 13-dimensionalen MFCC-Vektoren [3] extrahiert. Dabei betragen die Fensterbreite 25 ms und der Vorschub 10 ms. Der Frequenzbereich der zur Konvertierung auf die Mel-Skala genutzten Dreieckfilter wurde auf 0 bis 4 kHz begrenzt. Entlang aller 13

Dimensionen des Merkmalsraumes erfolgte ein Min/Max-Mapping des Wertebereichs auf ± 1 . Dies ermöglicht es, bei der folgenden Erstellung eines Vektorquantisierers die einfache euklidische Distanz als Distanzmaß zu nutzen, ohne dass sich die unterschiedlich großen Wertebereiche der einzelnen Dimensionen in den MFCC-Merkmalen negativ auf das Codebuch auswirken. Um bei der Codebuchgenerierung Unabhängigkeit von der ungleichen A-priori-Verteilung der Proben zu erlangen, wurden für beide Klassen separat mit dem Linde-Buzo-Gray-Algorithmus Codebücher mit je 2048 Einträgen trainiert, die anschließend zu einem Codebuch mit 4096 Einträgen zusammengefasst wurden. Durch die Anwendung dieses Quantisierers auf alle Trainingsdaten wurden die 13-dimensionalen Features für die weiteren Trainingsschritte als eindimensionale Codebuchindices repräsentiert. Auf dieser Basis wurden zwei diskrete HMM für Schnarchen bzw. Nebengeräusche trainiert, wobei das Codebuch den möglichen Emissionen Ξ (nach [3]) entspricht. Die Anzahl der Zustände wurde zum Modellieren des Schnarchens (siehe Abb. 1) auf fünf festgelegt, ebenso für die Nebengeräusche. Das Training erfolgte durch den Baum-Welch-Algorithmus. Nach dem Training wurde die Anpassung der HMM wie in Abb. 2 gezeigt überprüft.

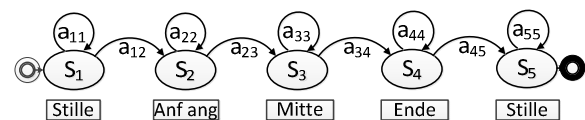


Abbildung 1: Modellierung eines Schnarchgeräusches mit einem Hidden-Markov-Modell mit fünf Zuständen. Die initiale Schätzung der Übergangswahrscheinlichkeiten für den Trainingsvorgang gibt die Struktur des HMM vor.

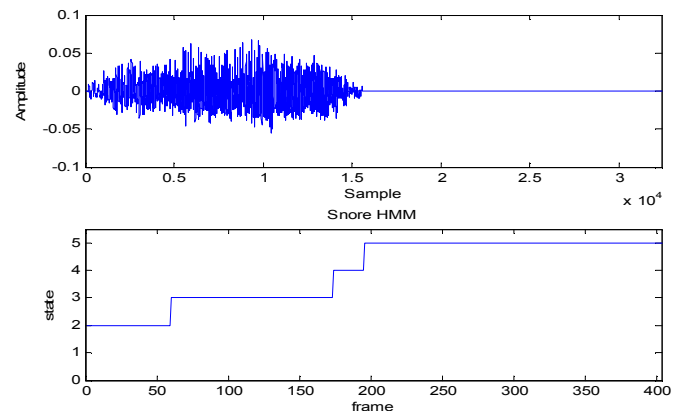


Abbildung 2: Oben: Abtastwerte eines Schnarchsignals. Zu erkennen ist, dass das Schnarchen zunächst zu seiner vollen Ausprägung anschwillt, diese dann eine Zeit lang beibehält und anschließend wieder abklingt. Unten: wahrscheinlichste Zustandsfolge des HMM für Schnarchsignale. Diese spiegelt die Realität wie gewünscht wieder.

Um die Analyse von fortlaufenden Audiosignalen zu ermöglichen, werden die beiden HMM miteinander verknüpft. Das in Abb. 3 gezeigte System entsteht.

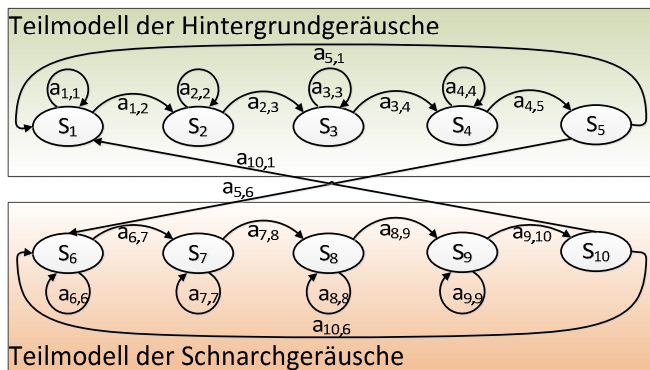


Abbildung 3: Die beiden HMM werden durch die Ergänzung von Übergängen verbunden. Die Emissionswahrscheinlichkeiten werden für jeden Zustand direkt aus dem zuvor trainiertem HMM übernommen. Die Übergangswahrscheinlichkeiten werden entsprechend Tabelle 1 modifiziert.

		Zielzustand									
		S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇	S ₈	S ₉	S ₁₀
Ausgangszustand	S ₁	a _{1,1}	a _{1,2}	0	0	0	0	0	0	0	0
	S ₂	0	a _{2,2}	a _{2,3}	0	0	0	0	0	0	0
	S ₃	0	0	a _{3,3}	a _{3,4}	0	0	0	0	0	0
	S ₄	0	0	0	a _{4,4}	a _{4,5}	0	0	0	0	0
	S ₅	a _{5,1}	0	0	0	0	a _{5,6}	0	0	0	0
	S ₆	0	0	0	0	0	a _{6,6}	a _{6,7}	0	0	0
	S ₇	0	0	0	0	0	0	a _{7,7}	a _{7,8}	0	0
	S ₈	0	0	0	0	0	0	0	a _{8,8}	a _{8,9}	0
	S ₉	0	0	0	0	0	0	0	0	a _{9,9}	a _{9,10}
	S ₁₀	a _{10,1}	0	0	0	0	a _{10,6}	0	0	0	0

Tabelle 1: Übergangswahrscheinlichkeiten des HMM-Systems für fortlaufende Audiospuren. Die Übergänge a_{5,1}, a_{5,6}, a_{10,1} und a_{10,6} wurden manuell eingebracht. Der rote Bereich stammt aus dem Modell für Schnarchen und der grüne Bereich aus dem Modell für Nebengeräusche.

Experimente und Resultate: Klassifikation

Die Klassifikation von unbekanntem Geräuschen O kann nun anhand ihrer Erzeugungswahrscheinlichkeit $P(O|\lambda)$ [3] aus den beiden trainierten HMM λ durchgeführt werden. Dazu wird die Erzeugungswahrscheinlichkeit des Schnarch-HMM $P(O|\lambda_{Schnarchen})$ mit der des Modells für Nebengeräusche $P(O|\lambda_{Nebengeräusche})$ verglichen. Somit wird ermittelt, bei welchem Modell die Produktion der Emissionsfolge O wahrscheinlicher ist. Mittels dieses Verfahrens konnte im Test gegen Validierungsdaten (10% des Korpus, vor dem Training zurückgehalten) eine Genauigkeit der Klassifikation von 96,05 % erreicht werden.

Tabelle 2: Confusion Matrix der Klassifikation

		Detektierte Klasse	
		Schnarchen	Nebengeräusche
Wirkliche Klasse	Schnarchen	135 TP 96,43 %	5 FN 3,57 %
	Nebengeräusche	2 FP 5,41 %	35 TN 94,59 %

Erkennung in fortlaufenden Audiosignalen

Die Analyse von fortlaufenden Audiosignalen O kann durch Viterbi-Dekodierung des HMM-Modells in Abb. 3 erfolgen. So wird die wahrscheinlichste Folge der verborgenen Zustände q^* ermittelt, die die zu prüfende Emissionsfolge O erzeugt [3]. Sobald ein Zustand, der zum Modell des Schnarchens gehört (S₆, S₇, S₈, S₉, siehe Abb. 3), am wahrscheinlichsten ist, wird vom System ein Schnarchen erkannt. Die Position dieser Zustände in der Emissionsfolge entspricht der Position der Frames mit dem detektierten Schnarchen in der zugrundeliegenden Audiospur.

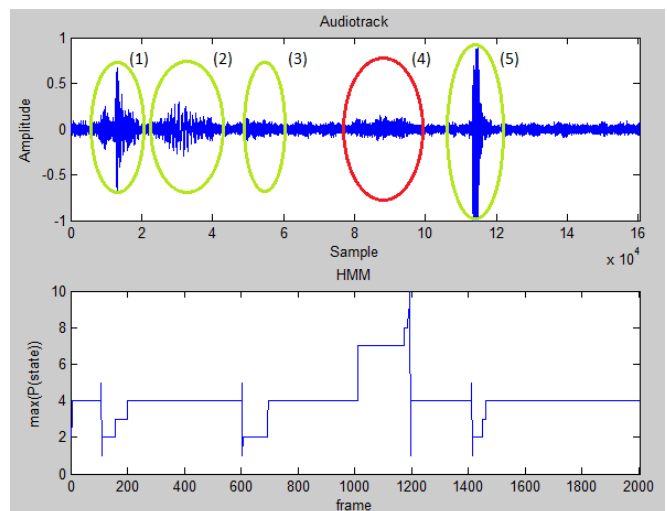


Abbildung 4: Erkennung von Schnarchen und Nebengeräuschen in einer Signalfolge von 20 Sekunden Länge. Dargestellt sind Signal im Zeitbereich (oben), Ergebnis der Viterbi-Decodierung (unten); (1) Husten, (2) Bettdeckengeräusch, (3) lautes Atmen, (4) Schnarchen und (5) Räuspern und Sprache.

Fazit

Es wurde ein System entwickelt, das mit Hilfe gebräuchlicher Methoden der Sprachverarbeitung Schnarchgeräusche mit einer Genauigkeit der Klassifikation von 96 % erkennen und in fortlaufenden Audiosignalen orten kann. Wegen der durchgehenden aufwandsarmen Realisierung ist die Eignung des Lösungsvorschlags für die Anwendung in Echtzeit auf Embedded Hardware höchst wahrscheinlich. Auch ist anzunehmen, dass durch eine Vergrößerung des Trainingskorpus eine Verbesserung der Generalisierungsfähigkeit zu erreichen ist.

Literatur

- [1] Mousa, M.; Feldes, S.; Krauß, K.-H.: Eingebettetes System zur Schnarch-Erkennung und Schnarch-Unterbindung; URL: hs-mannheim.de/forschung-und-entwicklung.html, FuE-Profil 2013
- [2] Duckitt, W.: Automatic detection, segmentation and assessment of snoring from ambient acoustic data. *Physiol. Meas.*, vol. 27, no. 10 (2006), 1047–1056
- [3] E. Schukat-Talamazzini; Automatische Spracherkennung - Grundlagen, statistische Modelle und effiziente Algorithmen; Vieweg, Braunschweig, 1995