Aus dem Institut für Technologie und Biosystemtechnik

Gerhard Jahns

Automatischer Ruferkenner für landwirtschaftliche Nutztiere - Tierstimmerkennung

Veröffentlicht in: Landbauforschung Völkenrode ; 56(2006)1-2:31-37

Braunschweig
Bundesforschungsanstalt für Landwirtschaft (FAL)
2006

Automatischer Ruferkenner für landwirtschaftliche Nutztiere – Tierstimmerkennung

Gerhard Jahns*

Zuammenfassung

Ziel automatischer Ruferkenner[1] ist die effiziente und artgerechte Überwachung landwirtschaftlicher Nutztiere zur Erhaltung und Verbesserung der Gesundheit und des Wohlbefindens der Tiere. Die Natur hat im Laufe der Evolution auch bei Tieren Formen der akustischen Kommunikation entwickelt. Landwirte und Ethologen sind davon überzeugt, dass die von den Tieren selbst ausgehenden Lautäußerungen wertvolle Informationen über ihr Befinden und ihren Zustand liefern. Die Anatomie und die physiologischen Vorgänge der Lauterzeugung bei Wirbeltieren sind evolutionsbedingt sehr ähnlich. Es ist daher naheliegend, bei der Entwicklung automatischer Ruferkenner für landwirtschaftliche Nutztiere auf Erfahrungen und bewährte Methoden der menschlichen Spracherkennung zurückzugreifen. Die abschließend wiedergegebenen Ergebnisse beruhen auf dem Einsatz von Hidden-Markov-Modellen und einer Parametrisierung der Lautäußerungen mittels Mel-Cepstral-Koeffizienten. Für den Erfolg eines Ruferkenners kommt, neben den eingesetzten Methoden, dem verwendeten Datenkorpus eine entscheidende Bedeutung zu.

Schlüsselworte: Bioakustik, Digitale Signalverarbeitung, Hidden-Markov-Modelle (HMM), Landwirtschaft, Tiergesundheit, Tierschutz, Wirtschaftlichkeit, Ruferkenner, Schallanalyse

Abstract

Automatic identification of farm animal utterances - animal call-recognition

Acoustic monitoring of farm animals may serve as an efficient and species-appropriate management tool for enhancing animal health, welfare, and farm efficiency. In the course of evolution, nature developed manifold means of communication. Sound is one of the most important to convey information and to express emotional states and conditions. Farmers and ethologists are convinced that animal utterances provide valuable information about the animals' state-of-being and condition. Despite the complexity of human speech and the size of the human vocabulary , which is unique in the animal realm, the production and reception of sounds in vertebrates have much in common with human processes. This encourages science to adapt methods and experiences from speech-recognition to recognise animal calls. The problem of animal independent call-recognition is comparable to speaker independent word spotting in speech-recognition. In speech-recognition, double stochastic processes such as Hidden Markov Models (HMMs) have proved very efficient. They are applied here to recognise animal calls, using utterances of cows as an example. Results presented here are based on the use of Hidden-Morkov-Models and features such as Mel-Frequency-Cepstral Coefficients. Beside the methods applied, the success of a call- recogniser very much depends on a representative and comprehensive data corpus. The results reveal that HMMs are well suited for animal call-recognition.

Keywords: Call-recogniser, animal vocalization, Hidden Markov Model (HMM), sound analysis, bio acoustics, digital signal processing, farming, animal health, animal protection.

* Institut für Technologie und Biosystemtechnik, Bundesforschungsanstalt für Landwirtschaft (FAL), Bundesallee 50, 38116 Braunschweig; E-Mail: sigi.gerd.jahns@gmx.de

[1] Die Bezeichnung Ruferkenner wurde in Anlehnung an den Begriff Spracherkenner gewählt.

1 Motivation und Zielsetzung

Eine Überwachung landwirtschaftlicher Nutztiere zur Erhaltung und Verbesserung der Tiergesundheit und des Wohlbefindens sind Voraussetzung für eine artgerechte und effiziente Tierhaltung sowie für den Verbraucherschutz und die Lebensmittelsicherheit. Es ist generell wenig sinnvoll, Qualität am Ende einer Produktionskette "erprüfen" zu wollen. Vorzuziehen ist stattdessen eine kontinuierliche prozessbegleitende Überwachung und Qualitätskontrolle. Dabei kommt vor allem berührungslosen, nicht invasiven und zerstörungsfreien Verfahren eine besondere Bedeutung zu, die das natürliche Verhalten der Tiere nicht beeinflussen oder gar einschränken, neben visuellen und olfaktometrischen Verfahren sind dies in erster Linie akustische Verfahren.

Während Landwirte in kleinen Familienbetrieben ihre Tiere noch "persönlich" kennen, ist dies in Großbetrieben, bei denen die Versorgung der Tiere u. U. auch durch Nichtlandwirte erfolgt, nicht mehr der Fall. Seit langem ist man daher bemüht, individuelle Merkmale und Parameter der Tiere zu messen. Das Schlagwort hierfür lautet "precision livestock farming". In der Milchviehhaltung werden beispielsweise Temperatur, Leitfähigkeit und pH-Wert der Milch, Lebendgewicht, Aktivität und Futteraufnahme der Tiere gemessen, um die individuellen Bedürfnisse und die Gesundheit der Tiere zu überwachen und die Tiere artgerecht zu versorgen. Derartige Systeme sind nicht nur teuer, sie können darüber hinaus das natürliche Verhalten der Tiere beeinflussen, außerdem arbeiten sie absätzig, d. h. die Erfassung der Daten erfolgt nur zu bestimmten Zeiten des Tages, z.B. während des Melkens.

Andererseits haben die Tiere im Laufe ihrer Evolution Formen der Kommunikation entwickelt, um Informationen auszutauschen (Hauser, M.D., 1996). Die akustische Kommunikation nimmt dabei einen besonderen Stellenwert ein. Landwirte und Ethologen sind davon überzeugt, dass die von den Tieren selbst ausgehenden Lautäußerungen wertvolle Informationen über ihren Zustand und ihr Befinden liefern können. Eine akustische Überwachung lässt sich zudem berührungslos und kostengünstig realisieren. Die Überwachung kann ohne Beeinträchtigung der Tiere 24 Stunden am Tag erfolgen, wobei sich eine größere Zahl von Tieren gleichzeitig überwachen lässt. Bei einer entsprechenden Erweiterung der akustischen Analyse ist darüber hinaus auch ohne wesentlichen Mehraufwand eine Überwachung technischer Einrichtungen[2] in einem Stall möglich.

Ziel der nachfolgend geschilderten Arbeiten ist ein Ruferkennungssystem, d.h. ein System, das in der Lage ist, Lautäußerungen von Tieren zu unterscheiden und ihrer Bedeutung nach zu erkennen. Ist der erkannte Ruf von Bedeutung für den Landwirt, so kann dieser beispielsweise durch eine Meldung auf einem Bildschirm seines Computers oder ein akustisches Signal informiert oder alarmiert werden. Auch die Auslösung von Funktionen ist möglich. Bei einem für die Praxis geeigneten System würde der Landwirt von den im nachfolgenden aufgezeigten und diskutierten Details natürlich nichts erfahren; sie wären in dem Programm verborgen. Die bereits bestehenden Mess- und Überwachungssysteme des Betriebes ließen sich durch eine solche akustische Überwachung wirkungsvoll ergänzen. Für den Aufbau eines Ruferkenners ist es erforderlich, das Repertoire und die Bedeutung der Rufe einer Tierart, im Nachfolgenden sind es Kühe, zu kennen. Mittels einer möglichst großen Sammlung bekannter Rufe werden mathematische Modelle der Rufe gebildet, denen dann die unbekannten Rufe zugeordnet werden können. Vergleichbar ist diese Aufgabe mit der eines sprecherunabhängigen Spracherkennungssystems für eine fremde Sprache, einschließlich einem Übersetzer.

2 Datenkorpus

Für den Aufbau eines Ruferkenners ist eine repräsentative und umfassende Sammlung von Rufen (Datenkorpus) der jeweiligen Tierart erforderlich, deren Bedeutung bekannt sein muss. Ein solcher Datenkorpus muss möglichst groß sein. Damit die Ruferkennung tierunabhängig ist, müssen die Aufnahmen von möglichst vielen Tieren stammen. Die Zahl der Rufe pro Tier und pro Bedeutung sollte zehn, besser hundert, nicht unterschreiten. Je nach Größe des Rufrepertoires der betreffenden Tierart sollte der Datenkorpus einige hundert bis tausend Rufe enthalten. Von den zur Verfügung stehenden Rufen werden üblicherweise ca. 70 % für das sog. Training des Systems und 30 % für dessen Verifikation verwendet. Für jede Tierart ist im Vorfeld stets zu prüfen, ob die Abtastfrequenz[3] (sampling rate), mit der die Aufnahmen digitalisiert werden, ausreichend hoch ist (Nyquistfrequenz). Für Kühe, die im Folgenden als Beispiel dienen, reicht eine Abtastfrequenz von 11025 Hz aus, da alle wesentlichen Merkmale der Rufe unter 5000 Hz liegen.

3 Aufnahmebedingungen

Die Erstellung des Datenkorpus wird sich besonders bei Großtieren schon aus Kostengründen kaum in besonders

[2] Verfahren dieser Art sind in der Industrie Stand der Technik

[3] Bei der Aufnahme eines Schallereignisses mit einem Mikrophon erhält man einen kontinuierlichen Spannungsverlauf f(t), der den Schalldruckschwankungen proportional ist. Die Signalverarbeitung erfolgt heute nahezu ausschließlich digital, zu diesem Zweck wird ein solch kontinuierlicher Signalverlauf an diskreten Stellen abgetastet, wodurch Information verloren geht. Es lässt sich zeigen, dass eine kontinuierliche Funktion f(t) aus den Abtastwerten $f_n = f(nT)$ vollständig rekonstruierbar ist ($f_A = 1/T \geq 2f_G$), wenn die Abtastfrequenz (f_A) oberhalb der doppelten Grenzfrequenz (f_G), der sog. Nyquistfrequenz, liegt.

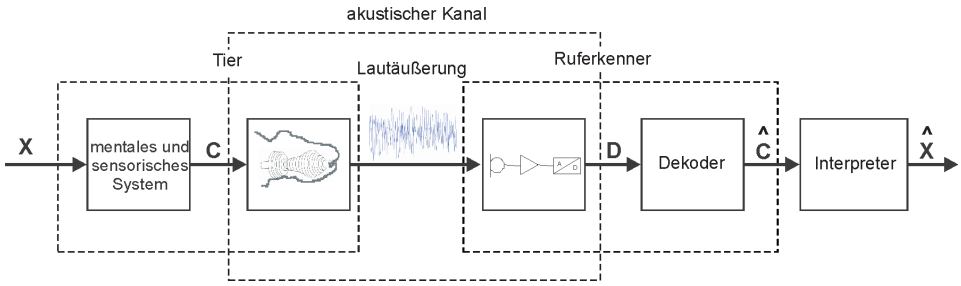

Abb. 1:
Quelle-Kanal-Erkenner Modell

X = Befinden/Zustand des Tieres
C = Ruffolge
D = Zahlenfolge
$\hat{C}$ = Rufäquivalent, das die Ruffolge am wahrscheinlichsten wiedergibt
$\hat{X}$ = Text, der das Befinden bzw. den Zustand des Tieres am wahrscheinlichsten wiedergibt

gestalteten akustischen Räumen durchführen lassen, sondern nur unter realen Bedingungen. Folglich sind derartige Aufnahmen zwangsläufig mit Störungen behaftet, als da sind: Echo, Dämpfung, Absorption, Resonanz usw. Dies ist ein weiterer Grund warum die Zahl der Rufe des Datenkorpus möglichst groß sein muss und hat auch Konsequenzen für die Parameterauswahl (s.u.). Natürlich erschwert dies die Entwicklung eines Ruferkenners, es ist aber in sofern nicht nachteilig, weil der Ruferkenner später in der Praxis in eben einer solchen realen Umgebung eingesetzt werden soll. Im Gegensatz zur Spracherkennung ist es nicht möglich, Tiere zu Lautäußerungen einer bestimmten Art aufzufordern. So bleibt nur die Möglichkeit, dass Experten die die Bedeutung der Rufe kennen, diese richtig zuordnen. Ist dieses Wissen nicht vorhanden, ist es erforderlich für die Tiere entsprechende Bedingungen und Zustände zu schaffen, um sie so zu definierten Lautäußerungen zu veranlassen. Ein sehr aufwendiges Vorgehen, bei dem ebenfalls die Kenntnisse von Ethologen gefragt sind.

4 Quelle – Kanal – Ruferkenner

Die Anatomie und die physiologischen Vorgänge der Lauterzeugung bei Wirbeltieren sind evolutionsbedingt sehr ähnlich. Diese Ähnlichkeit ist um so größer, je näher die Arten miteinander verwandt sind. Das Schema der Erzeugung von Lautäußerungen, ihre Erkennung und Übersetzung (Interpretation) ist in der Abb. 1 dargestellt. Aufgrund eines Zustandes oder einer Empfindung wird vom Zentralnervensystem das Signal zur Erzeugung eines Rufes oder einer Rufsequenz generiert. Die physikalische Erzeugung der Lautäußerung erfolgt im Vokaltrakt. Dabei strömt bei Wirbeltieren Luft aus der Lunge durch den Kehlkopf (Larynx) und regt dabei die Stimmbänder (Glottis) zu periodischen Schwingungen an. Die Frequenz dieser Schwingungen wird durch die Tension der Kehlkopfmuskeln bestimmt. Im nachfolgenden Vokaltrakt erfolgt eine weitere Formung der Lautäußerungen und schließlich die Abstrahlung von den Lippen oder den Nasenöffnungen. Stimmlose Lautäußerungen unterscheiden sich von stimmhaften darin, dass an den Stimmbändern keine periodischen Schwingungen (Grundfrequenz) erzeugt werden, sondern ein weißes Rauschen. Das Ergebnis sind in jedem Fall Schalldruckschwankungen. Als Schall werden Druck- und Dichteschwankungen in einem elastischen Medium bezeichnet, die sich in Form von Längs- bzw. Longitudinalwellen ausbreiten, die sich in der Luft fortpflanzen.

Abb. 2 zeigt das hydraulische Modell und Abb. 3 das mathematische Modell dieses Vorganges. Trotz prinzipieller Gleichheit weist die Spracherzeugung beim Menschen einige Besonderheiten auf (Paulsen, K., 1967). Beim Menschen ist der Kehlkopf zur Lunge hin verlagert, wodurch sich gegenüber anderen Säugetieren ein deutlich größerer Rachenraum (Pharynx) ergib. Des Weiteren zweigen Nasen- und Mundhöhle nahezu rechtwinklig vom Rachenraum ab. Auch wenn Tiere Lautäußerungen

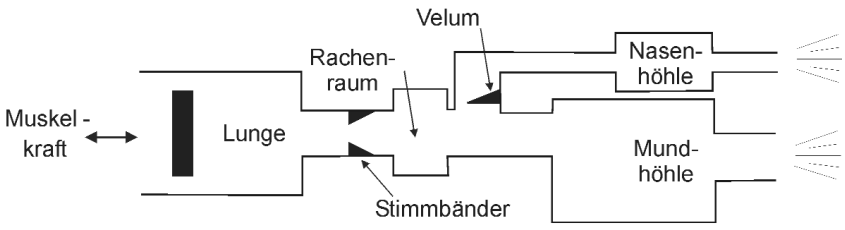

Abb. 2:
Schematische Darstellung eines Vokaltraktes bei Tieren

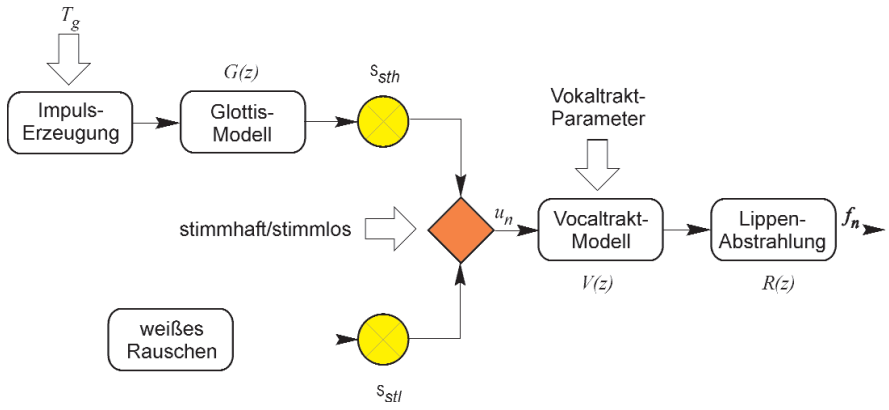

Abb. 3:
Quelle-Filter Modell nach G. Fant sowie das Faltungsprodukt bzw. die ztransformierte Darstellung des Signals

zur Kommunikation entwickelt haben, ist diese nach Art und Umfang weit von dem der menschlichen Sprache entfernt. Die Diskussion, wie die menschliche Sprache entstand bzw. was dafür Voraussetzung und erforderlich war, ist noch nicht abgeschlossen und soll hier auch nicht weiter verfolgt werden. Sicher ist, dass in Rudeln lebende und jagende Tiere ein umfangreicheres Repertoire von Lautäußerungen zur Kommunikation entwickelt haben als Beutetiere oder Tierarten, die als Einzelgänger leben.

Die von den Lippen oder der Nase abgestrahlten Schalldruckschwankungen werden vom Mikrophon in elektrische Spannungen umgesetzt, verstärkt und digitalisiert (Abb. 1). Der Frequenzgang kostengünstiger Elektretmikrophone und Verstärker sowie der meisten Soundkarten in PCs ist völlig ausreichend. Die Kosten für den erforderlichen Hardwareaufwand betragen heute wenige Euro und sind damit vernachlässigbar gering. Dies gilt allerdings nur, wenn die Lautäußerungen der betreffenden Tierart im Bereich zwischen 20 Hz und 20 kHz liegen (Jahns, G., 1998). Dies ist für jede Tierart zu prüfen. Aus den digitalisierten Schalldruckschwankungen wird dann mittels geeigneter Verfahren ein mathematisches Rufäquivalent ermittelt, das den ursprünglichen Ruf oder die Ruffolge wiedergibt.

5 Ruferkennung

Die Ruferkennung ist, wie die Spracherkennung (Schukat-Talamazzini, E.G., 1995; Deller, J.R. et al., 1987), ein Klassifikationsprozess. Aus einem Ruf werden bestimmte Merkmale, im vorliegenden Fall wurden Mel-Cepstral-Koeffizienten[4] verwendet, anhand derer der zu erkennende Ruf mit den bekannten Rufen des Rufrepertoires verglichen und dann dem zugeordnet wird, dem er am ähnlichsten ist. Ein besonderes Problem besteht darin, dass tierische wie menschliche Lautäußerungen hinsichtlich Frequenz, Energie und Dauer variieren und zwar auch dann, wenn sie die gleiche Bedeutung haben. Besonders

[4] Koeffizienten, die unter Berücksichtigung des psychoakustischen Tonhöhenempfindens (MEL-Skala) als inverse Fouriertransformierte des logarithmischen Leistungsdichtespektrums berechnet werden.

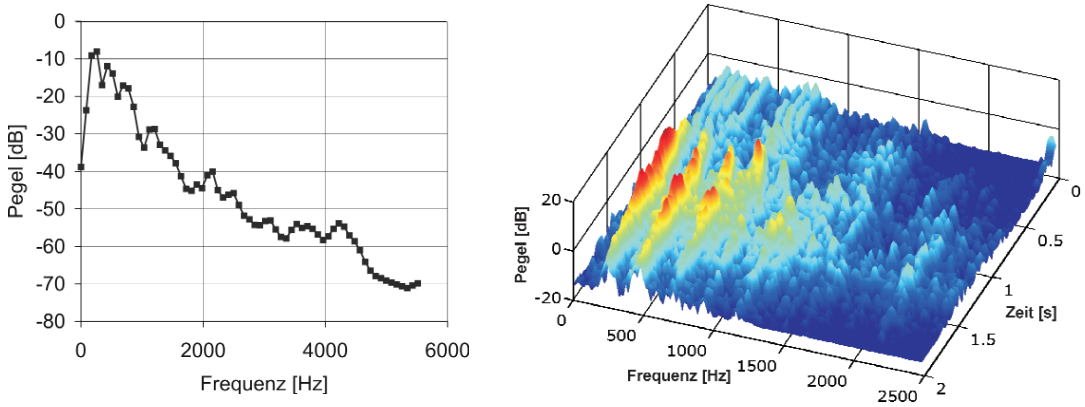

Abb. 4:
Energiedichtespektrum des Rufes einer Kuh wegen verspätetem Melken (links) und Spektrogramm desselben Rufes (rechts)

die zeitliche Dehnung, die selbst bei ein und demselben Individuum von Lautäußerung zu Lautäußerung und auch innerhalb einer Lautäußerung variieren kann, bereitet besondere Schwierigkeiten. Sie schließt den Einsatz vieler Klassifizierungsverfahren aus, zumindest wenn diese ohne entsprechende Vorverarbeitung, wie beispielsweise durch *time warping*[5], angewendet werden.

Generell kann man einen Ruf als eine Einheit behandeln. Dies hat den Vorteil, dass sich Merkmalsvektoren einfach gewinnen und klassifizieren lassen und das Problem der zeitlichen Dehnung entfällt. Es stößt aber auf Schwierigkeiten bei der praktischen Realisierung. Denn bei Rufen, die in realer Umgebung aufgenommen werden, lassen sich Anfang und Ende nur schwer automatisch bestimmen. Auch das automatische Erkennen überlappender Rufe bereitet Schwierigkeiten. Das sichere Erkennen von Anfang und Ende einer Lautäußerung hat aber entscheidenden Einfluss auf die Fehlerrate bei der Erkennung (Wilpon, J.G. et al., 1984). In früheren Untersuchungen konnte gezeigt werden, dass sich einzelne Tiere an Hand der Energiedichtespektren ihrer Rufe recht gut unterscheiden lassen. Die Unterscheidung von Rufen unterschiedlicher Bedeutung an Hand des Energiedichtespektrums war dagegen unzureichend (Jahns, G. et al., 1997).

Aus der Abb. 4 wird aber auch deutlich, dass bei Verwendung des Energiedichtespektrums wertvolle Informationen, insbesondere die der zeitlichen Änderung von Merkmalen innerhalb eines Rufes verloren gehen. In der Abb. 4 ist links das Energiedichtespektrum[6] des gesamten Rufes einer Kuh aufgrund verspäteten Melkens und rechts das Spektrogramm[7] desselben Rufes dargestellt.

In der Spracherkennung gewinnt man Merkmalsvektoren einer Lautäußerung, indem man die in ein zeitliches Fenster (delta) fallenden Daten auswertet. Dieses Fenster muss einerseits so klein gewählt werden, dass der Signalverlauf darin als stationär gelten kann, andererseits so groß, dass alle Merkmale eindeutig und zuverlässig bestimmt werden können. Dieses Fenster wird dann um Zeitschritte (step), die kleiner als die Länge des Fensters sind, verschoben, dadurch ergibt sich eine Überlappung.

Die Abb. 5 zeigt den zeitlichen Verlauf der Amplitude eines Rufes (rechts oben) mit dem Fenster und darunter das Spektrogramm mit gleicher Zeitachse und der zeitlichen Verschiebung des Fensters. Links im Bild sind für den gefensterten Signalabschnitt die Energiedichte und die lineare Vorhersage über der Frequenz dargestellt. Die Anzahl der möglichen Parameter zur Beschreibung eines Rufes ist außerordentlich groß. Ihre geschickte Wahl hat auf den Erfolg und Rechenaufwand des Verfahrens unmittelbaren Einfluss.

Da die Ruferkennung tierunabhängig sein soll, wird man Cepstral-Koeffizienten gegenüber linearen Vorhersagekoeffizienten bevorzugen. Linear Prediction Coefficients[8] (LPC) spiegeln vor allem die Eigenschaften des Vokaltraktes und damit des Individuums wieder, während Mel-Frequency Cepstral Coefficients (MFCC) geeigneter sind, eine vom Individuum unabhängige Ruferkennung zu

[5] Ein mathematisches Verfahren (dynamische Programmierung), um die unterschiedlichen zeitlichen Verzerrungen zweier gleichbedeutender Lautäußerungen zu kompensieren.

[6] Das Energiedichtespektrum (Power Spectral Density (PSD)) gibt den relativen Energieanteil eines Schallereignisses in Abhängigkeit von der Frequenz wieder.

[7] Das Spektrogramm gibt die relative Energiedichte eines Schallereignisses als Intensität (als Grautöne oder farblich) in Abhängigkeit von Frequenz (Ordinate) und Zeit (Abszisse) wieder.

[8] Koeffizienten zur Schätzung der aktuellen Werte eines Schallereignisses, die aus einer Linearkombination der vorangegangenen Werte berechnet werden.

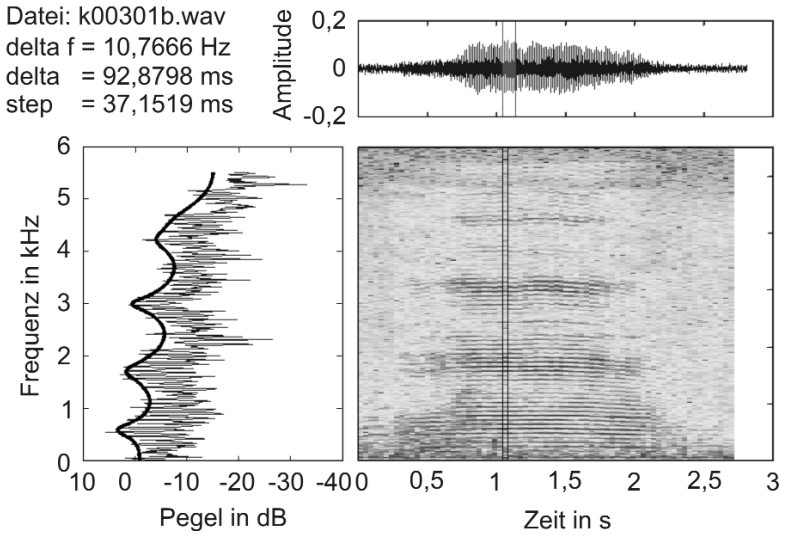

Abb. 5:
Oszillogramm, Spektrogramm, Energiedichtespektrum (PSD) und lineare Vorhersage (LP)

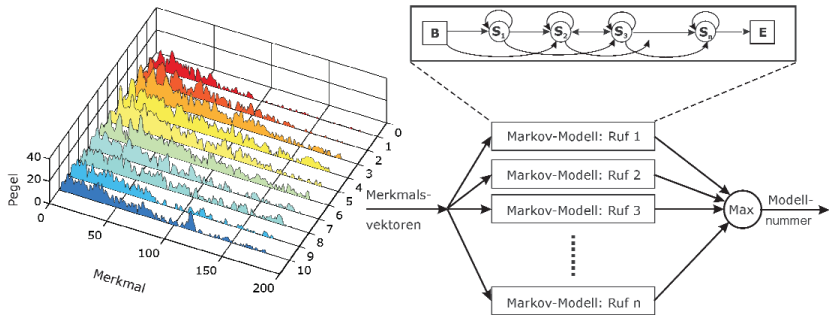

Abb. 6:
Ruferkennung mit Hidden-Markov-Modellen

ermöglichen. Im Hinblick auf den späteren Online-Einsatz des Ruferkenners unter realen Bedingungen, z. B. in einem Stall, ergeben sich weitere Einschränkungen bei der Wahl der Parameter. So ist beispielsweise der Absolutwert der Energie ungeeignet, da Rufrichtung und Entfernung zum Mikrophon wechseln und nicht bekannt sind.

6 Auswahl des Verfahrens

Bei der Ruferkennung besteht, wie bei der Worterkennung, die Aufgabe, die Identität der jeweiligen Lautäußerung durch Vergleich mit dem vorher zu erstellenden Rufrepertoire zu ermitteln. Dies kann einmal durch eine Abstandsklassifikation geschehen, bei der die Entscheidung auf der Minimierung des Abstandes der Eingabesequenz zu einer Referenzfolge beruht oder durch eine statistische Modellierung. Bei letzterer fällt die Entscheidung zu Gunsten desjenigen Rufes, der mit der größten Wahrscheinlichkeit dem zu erkennenden am ähnlichsten ist. Beide Strategien lassen sich für die Erkennung ganzer Rufe ebenso anwenden wie zur Erkennung phonetischer Segmente von Rufen.

Es sind keine Anzeichen dafür bekannt, dass die Lautäußerungen der Wirbeltiere nicht, ebenso wie die des Menschen, von unterschiedlicher variierender zeitlicher Dauer sind. Abhängig von dem Artikulationstempo und -rhytmus entfällt dadurch auf jedes einzelne phonetische Segment einer Lautäußerung eine nicht vorhersagbare Anzahl von Merkmalsvektoren. Jeder dieser Merkmalsvektoren selbst umfasst neben seinem phonetischen Gehalt individuen-, umgebungs- und verschleifungsbedingte Informationsanteile. Diese und die variable Dauer der Lautäußerungen erschweren ihre Identifikation.

In den siebziger Jahren versuchte man in der Spracherkennung dem Problem der nichtlinearen zeitlichen Verzerrung mittels Mustervergleich durch dynamische Zeitverzerrung (dynamic time warping) zu begegnen. Seit den 80er Jahren haben sich in der Spracherkennung statistische Modellierungen durch Markov-Modelle (HMM: Hidden-Markov-Model) bewährt. Sie haben u.a. den Vorteil, dass die Festlegung der Wort- und Phonemgrenzen bei den Trainingsdaten unkritisch ist, was bei anderen Verfahren aber zu erheblichen Fehlern führt (Wilpon, J.G., 1984). Dies ist ein wichtiger Gesichtspunkt bei Aufnahmen unter realen Einsatzbedingungen, bei denen Hintergrundgeräusche unvermeidlich sind.

Bei den Hidden-Markov-Modellen (Rabiner, L.R., 1989) handelt es sich um doppelt stochastische Prozesse mit N Zuständen und K Ausgängen, die je durch einen Parametersatz $\lambda = (\pi, A, B)$ definiert werden. Der N-dimensionale Vektor π legt die Wahrscheinlichkeiten der Anfangszustände fest. Die Übergangswahrscheinlichkeiten werden durch eine N x N dimensionale Parametermatrix A festgelegt. Die Ausgabeverteilung wird durch die N x K dimensionale Matrix B bestimmt. Die Abb. 6 zeigt ein solches links-rechts Modell. Die Matrix der Übergangswahrscheinlichkeiten A ist hier nicht voll besetzt. Die in denselben Zustand zurücklaufenden Schleifen modellieren die zeitliche Dehnung eines Phonsegments. Die einen oder auch mehrere Zustände überspringenden Schleifen modellieren Verschleifungen bzw. Auslassungen von Phonsegmenten. Zu jedem Zeittakt ist also ein Verbleiben im jeweiligen Zustand oder ein Übergang zu dem eines nachfolgenden Zustandes möglich. Dieser Übergang wird durch die Übergangswahrscheinlichkeiten (A) bestimmt. Jeder Zustand des Modells besitzt eine statistische Ausgabefunktion, die Symbole generiert, in diesem Fall Phonsegmente. Ein Beobachter kann lediglich die generierte Symbolfolge, nicht jedoch die Folge der inneren Zustände erkennen, daher die Bezeichnung „hidden“: Man unterscheidet des weiteren Markov Modelle mit diskreten Ausgabeverteilungen und solche mit kontinuierlichen Verteilungsdichten. Für die nachfolgend wiedergegebenen Ergebnisse wurden kontinuierliche Verteilungsdichten verwendet.

Generell kann man bei kleinem Rufrepertoire für jeden Ruf ein eigenes Modell erstellen. Bei großem Rufrepertoire, wie z. B. bei dem Wortschatz der menschlichen Sprache, bevorzugt man die Modellierung wiederkehrender Strukturen und aus diesen dann die Modellierung der jeweiligen Worte. Beide Verfahren sind dazu geeignet, aus kontinuierlichen akustischen Ereignissen einzelne Rufe zu erkennen (call spotting). Für die Ruferkennung der Lautäußerung von Kühen wurde, wegen der zu erwartenden

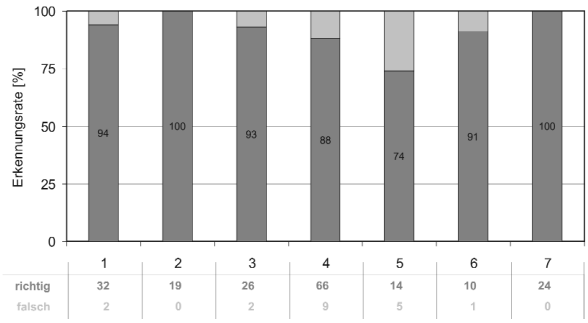

Abb. 7:
Erkennungsrate unterschiedlicher Lautäußerungen von Kühen mittels Hidden-Markov-Modellen

Art der Rufe:
1 Kuh ruft Kalb – temporäre Trennung
2 Kalb ruft Kuh – temporäre Trennung
3 Husten
4 Brunst
5 verspätetes Melken
6 geräuschvolles Einatmen
7 Hunger

geringen Zahl der unterschiedlichen Lautäußerungen, der Modellierung einzelner Rufe der Vorzug gegeben.

7 Ergebnis und Beurteilung

Die Abb. 7 gibt die mit dem derzeitigen System erzielten Erkennungsraten für 7 unterschiedliche Lautäußerungen wieder. Das Modell wurde mit insgesamt 478 Rufen trainiert. Für die Validation standen 210 Rufe zur Verfügung. Verwendet wurden Mel-Cepstral-Koeffizienten sowie deren Delta- und Beschleunigungskoeffizienten einschließlich der logarithmierten Energiewerte. Die Berechnung erfolgte mit dem Programmpaket HTK (Young, S. et al., 2000). Dass die Rufe der Kälber eindeutig von den anderen Rufen, die alle von erwachsenen Kühen stammen, unterschieden wurden, überrascht nicht. Die übrigen Ergebnisse sind durchaus ermutigend. Sie sollten m. E. an Hand einer breiteren Datenbasis und mit anderen Modellparametern noch eingehender untersucht werden. Hier ist vor allem die sachkundige Interpretation der Rufe durch Ethologen gefragt.

Literatur

Deller, J.R., J.G. Proakis JG, and J.H.L. Hansen JHL: (1987) Discrete-Ttime Pprocessing of speech signals. New York : MacmillanPrentice Hall 1987

Hauser, M.D.: (1996) The eEvolution of Ccommunication. Cambridge : MIT Press 1996

Jahns, G.: (1998) Understanding Aanimal Vvocalisation [online]. Available at < http://www.tb.fal.de/staff/jahns/animal.htm> 1998 [cited 18.01.2006]

Jahns, G., W. Kowalczyk W, and K. Walter. K (1997) An application of sound processing techniques for determining conditions of cows [online]. In: Proceeding of the 4th International Workshop on Systems, Signal and Image Processing, Poznan, 28.-30. May 1997. Available at < (PDF Dokument 94 kB http://www.tb.fal.de/staff/jahns/papers/pdf/posen.pdf> [cited 18.01.2006)

Paulsen, K.: (1967) Prinzipien der Stimmbildung in der Wirbeltierreihe und beim Menschen. Frankfurt a M : Akademische Verlagsgesellschaft 1967

Rabiner, L.R.: (1989) A Ttutorial on Hhidden Markov Mmodels and Sselected Aapplications in sSpeech Rrecognition. Proceedings of the IEEE, Vol. 77, No. (2,) February 1989

Schukat-Talamazzini, E.G.: (1995) Automatische Spracherkennung. Braunschweig : Vieweg 1995

Wilpon, J.G., L.R. Rabiner LR, and T.B. Martin TB. (1984) An improved word-detection algorithm for telephone-quality speech incorporating both syntactic and semantic constrains. AT&T Tech. J., Vol. 63, No. (3,): p 479-498, March 1984

Young, S., D. Kershaw D, J. Odell J, D. Ollason D, V. Valtchev V, and P. Woodland P. (2000) The HTK Bbook. Version 3.0 July 2000