Ereigniserkennung in akustischen Signalen Neugeborener

Olaf Hochmuth, Humboldt-Universität zu Berlin, Institut für Informatik

Einleitung

Im Rahmen eines geförderten Forschungsprojektes[1] ist die Entwicklung eines intelligenten ereignisgesteuerten Mikrofons zu untersuchen. Solch ein Mikrofon kann zum Beispiel zur Überwachung Neugeborener (Babyfon) eingesetzt werden. Im Gegensatz zu herkömmlichen Babyfonen sollen die Audiosignale digitalisiert und mit Hilfe der in modernen Wohnungen meist ohnehin vorhandenen Kommunikationswege übertragen werden. Dabei ist es sinnvoll, nicht permanent Audiosignale zu übertragen, sondern in Abhängigkeit sogenannter Ereignisse verschiedene Signale zu senden:

- Ruhe: keine Signalübertragung
- Geräusche: Audiosignalübertragung
- Gefahren: Audio- und Videosignalübertragung

Es ist also ein Klassifikator zu entwerfen, der typische Ereignisse erkennt. Die Vorgehensweise beim Klassifikatorentwurf war:

- Erhebung einer Lernstichprobe
- Merkmalsgewinnung und -auswahl
- Berechnung kanonischer Merkmale
- Reklassifikation der Lernstichprobe
- Übertragung der Klassifikationsregel auf einen DSP oder FPGA

Lernstichprobe

Für die Lernstichprobe standen befundete Audiosignale von 4 Neugeborenen mit insgesamt 29 Minuten Signaldauer zur Verfügung. Die Befundung durch die jungen Mütter reicht von ‚wach', ‚nuckelt' über ‚niest', ‚Schluckauf' bis zu ‚schreit' und ‚schimpft'. Beispielsweise gibt es für das Baby Friedrich 15 verschiedene Befunde.

Abb. 1: Aleksandr in zwei typischen Situationen (1 Tag alt)

Durch mehrmaliges Anhören bei gleichzeitigem Ansehen der Audiosignale (in Abb. 2 blau), werden diese händisch segmentiert. Es wird entschieden, welche Ausschnitte bzw. $regions\ of\ interest - ROI$ des Audiosignals zum Anlernen des Klassifikators geeignet sind (in Abb. 2 rot).

Merkmalsgewinnung

Bei der Beobachtung der aufgezeichneten Audiosignale wird schnell klar, dass Merkmale aus dem Zeitbereich und aus dem Frequenzbereich berechnet werden sollten.

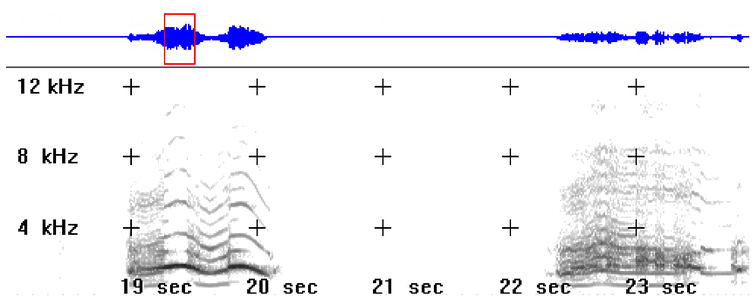

Abb. 2: Audiosignal und Spektrogramm von Hanna mit dem Befund ‚schreit'

Die Merkmale werden nur innerhalb der ROI gewonnen, und zwar in dem diese in möglichst viele, sich nicht überlappende Signalepisoden der Länge $N = 256$ zerlegt wird. Bei einer Abtastfrequenz f_A von $44{,}1\ \mathrm{kHz}$ beträgt die Episodendauer etwa $5{,}8\ \mathrm{ms}$. Es wird von jeder Episode ein Histogramm berechnet und daraus werden die Zeitbereichs-Merkmale bestimmt. Zu diesen 6 Merkmalen aus dem Zeitbereich gehören das quadratische Mittel, die Streuung, normierte Schiefe und Wölbung sowie die Entropie. Aus einer Episode kann die Anzahl der Nulldurchgänge bestimmt werden, hier ebenfalls ein Merkmal aus dem Zeitbereich. Aus jeder Episode wird auch jeweils ein DCT-Spektrum [3] berechnet. Aus dem zugehörigen Betragsspektrum werden dann die 3 Merkmale Grundfrequenz, Intensität der Grundfrequenz und Helligkeit [1] berechnet. Unter der ‚Helligkeit' eines Tones soll hier das gewogene Mittel über alle Frequenzen $n\ \Delta f$ im DCT-Betragsspektrum $\boldsymbol{DCT}$ verstanden werden, wobei das Gewicht die jeweilige Höhe DCT_n der Spektrallinie n ist:

$$\text{Helligkeit} = \sum_{n=0}^{N-1} DCT_n \cdot n\ \Delta f$$

Die spektrale Auflösung des DCT-Spektrums ist hier $\Delta f = \frac{f_A}{2N} = 86\ \mathrm{Hz}$. Alle 10 Merkmale werden in einen Merkmalsvektor geschrieben. Dieser stellt gemeinsam mit dem jeweiligen Befund und dem zeitlichen Beginn einer Episode eine Beobachtung im Sinne des Statistikprogrammes SAS® dar [2].

Merkmalsauswahl

Dem Statistikprogramm SAS® werden insgesamt 7671 Beobachtungen mit 15 Befunden übergeben. Da die Korrektklassifikationsrate für ein derart komplexes 15-Klassen-Problem erwartungsgemäß schlecht ist, wird versucht, die Klassenanzahl zu reduzieren. Durch Auswertung einer Matrix, die alle Korrektklassifikationsraten bei paarweiser Klassifikation enthält, werden 6 Klassen entfernt, die Korrektklassifikationsraten kleiner als 75% aufweisen.

…	schreit	trinkt	wach	weint	…
munter	92%	100%	**70%**	87%	…
niest	100%	100%	96%	84%	…
nörgelt	96%	93%	97%	**70%**	…
nuckelt	98%	98%	96%	92%	…
⋮	⋮	⋮	⋮	⋮	⋱

Für das Beispiel der beiden entfernten Klassen ‚munter' und ‚wach' zeigt Abb. 3 die hoffnungslose Situation im Merkmalsraum. Der Grund ist, dass schon die beiden Befunde schwer zu differenzieren sind. Neben diesen beiden Klassen betrifft das noch die Klassen ‚nörgelt' und ‚weint' sowie ‚langweilig' und ‚sauer'.

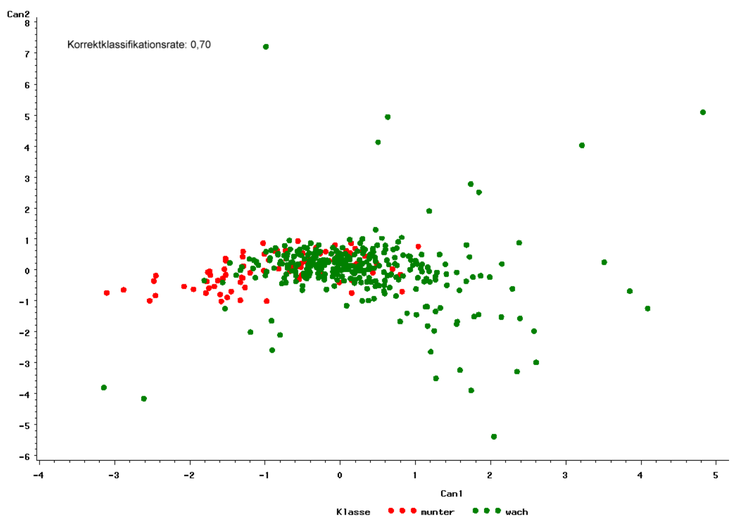

Abb. 3: Kanonischer Merkmalsraum mit 417 Beobachtungen der Klassen ‚munter' und ‚wach'

Außerdem werden 4 Klassen mit Klassenbesetzungszahlen kleiner als 50 aus der Lernstichprobe entfernt. Das betrifft die Klassen ‚niest', ‚nuckelt', ‚Schluckauf' und ‚trinkt'.
Für die verbleibenden 5 Klassen gibt die folgende Tabelle die Korrektklassifikationsraten bei Reklassifikation der Lernstichprobe an:

#	Klassen	
5	erzählt, fröhlich, lacht, schimpft, schreit	84%
4	erzählt, lacht, schimpft, schreit	91%
3	erzählt, schimpft, schreit	96%
2	erzählt, schreit	99%

Für alle Zeilen in der Tabelle reduziert SAS® die Merkmalsanzahl von 10 auf 7: das quadratische Mittel, Streuung, normierte Wölbung, Entropie, Intensität der Grundfrequenz, Helligkeit und Anzahl der Nulldurchgänge.

Klassifikationsregel im DSP oder FPGA

Die von SAS® exportierten Bildungsvorschriften $\boldsymbol{g_1}$ und $\boldsymbol{g_2}$ zur Berechnung der kanonischen Merkmale can_1 und can_2 ermöglichen die sensornahe Klassifikation in einem DSP oder FPGA. Für die Unterscheidung der Klassen ‚erzählt' und ‚schreit' gilt zum Beispiel:

$$\boldsymbol{g_1} = \left[1{,}8645\ \ -0{,}0304\ \ 0{,}0976\ \ -2{,}6237\ \ -0{,}0048\ \ -0{,}00098\ \ 0{,}0369\right]^{\mathsf{T}}$$

$$\boldsymbol{g_2} = \left[-0{,}9208\ \ -0{,}0036\ \ 0{,}1421\ \ -0{,}9204\ \ 0{,}09704\ \ 0{,}0011\ \ -0{,}0371\right]^{\mathsf{T}}$$

Für diese sensornahe Klassifikation sind fortlaufend für Episoden der Länge $N = 256$ alle 7 Merkmale zu berechnen und jeweils mit $\boldsymbol{g_1}$ und $\boldsymbol{g_2}$ zu wichten:

$$can_1 = \boldsymbol{g_1}^{\mathsf{T}} \cdot \left[\text{qMitt Streu nWölb Entrop GfInt Hell Zeros}\right]^{\mathsf{T}}$$

$$can_2 = \boldsymbol{g_2}^{\mathsf{T}} \cdot \left[\text{qMitt Streu nWölb Entrop GfInt Hell Zeros}\right]^{\mathsf{T}}$$

Abb. 4 zeigt einen Klassifikatortest mit dem Audiosignal ‚Friedrich erzählt'. Der Zeitraum $10{,}4\ \mathrm{s}$ bis $10{,}6\ \mathrm{s}$ wird in 34 Episoden zerlegt, es werden jeweils die 7 Merkmale und daraus die kanonischen Merkmale can_1 und can_2 berechnet. Das führt zu den 34 Testvektoren im kanonischen Merkmalsraum. Abb. 5 zeigt dasselbe für die Zerlegung des Zeitraums $12{,}2\ \mathrm{s}$ bis $15{,}8\ \mathrm{s}$ in 620 Episoden des Audiosignals ‚Friedrich schreit'.

Abb. 4: Kanonischer Merkmalsraum mit Trajektorie aus 34 Testvektoren des Signals ‚Friedrich erzählt'

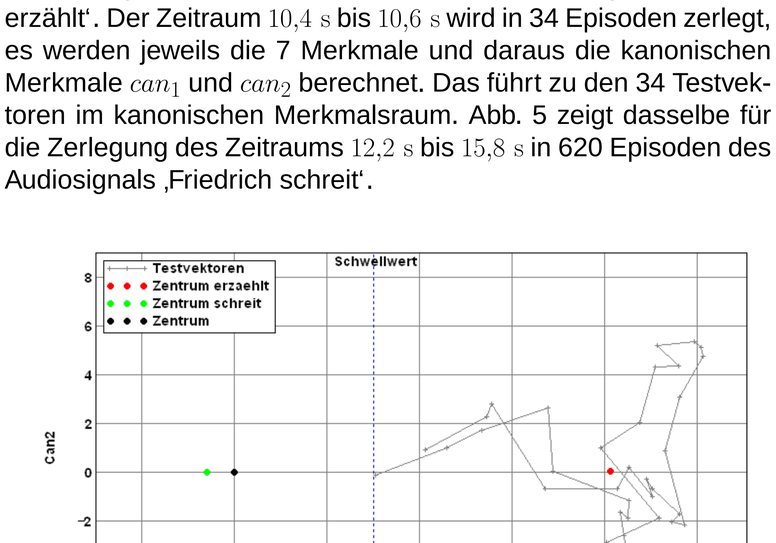

Abb. 5: Kanonischer Merkmalsraum mit Trajektorie aus 620 Testvektoren des Signals ‚Friedrich schreit'

Für das 2-Klassen-Problem ist das kanonische Merkmal can_1 ausreichend. Somit genügt es, im DSP oder FPGA über das Skalarprodukt von $\boldsymbol{g_1}$ und den 7 Merkmalen der aktuellen Episode das Merkmal can_1 zu berechnen. Anschließend ist lediglich mit einem Schwellwert (hier 3) zu vergleichen. Ist $can_1 \geq 3$, dann liegt die Klasse ‚erzählt' vor, andernfalls die Klasse ‚schreit'. Für einen Klassifikator mit Rückweisung kann mit weiteren Schwellwertvergleichen ein einfacher Quaderklassifikator realisiert werden.

Ergebnisse

Die Untersuchungen zeigen, dass es möglich ist, mit einfachen Mitteln der Signalverarbeitung und Mustererkennung ein intelligentes ereignisgesteuertes Mikrofon zu entwickeln. Die Signalverarbeitungsaufgaben im DSP oder FPGA sind einfach:

- Histogrammberechnung einer Episode
- Berechnung einiger Momente aus dem Histogramm
- DCT einer Episode
- Berechnung einiger Merkmale aus dem DCT-Spektrum
- Skalarprodukt von $\boldsymbol{g_1}$ und den 7 Merkmalen der Episode
- Vergleich mit einem Schwellwert

Alle Operationen sind in Echtzeit realisierbar.

Literatur

[1] Kosellek, D.: *Algorithmen zur Erkennung arttypischer Lautäußerungen von Vögeln.* Humboldt-Universität zu Berlin, Studienarbeit am Institut für Informatik, Berlin: Juni 2007

[2] Krämer, W.; Schoffler, O.; Tschiersch, L.: *Datenanalyse mit SAS®.* Berlin: Springer-Verlag 2005

[3] Meffert, B.; Hochmuth, O.: *Werkzeuge der Signalverarbeitung.* München: Pearson Studium 2004

Stand: 26. November 2010

[1] Prädiktionscodierung für hochwertige Audiosignale und intelligente Ereigniserkennung, Förderprogramm: Zentrales Innovationsprogramm Mittelstand – BMWi