EIN ECHTZEITFÄHIGES VERFAHREN ZUR ERKENNUNG VON BABYSCHREIEN

Renán Véliz, Stefan Feldes

Institut für Digitale Signalverarbeitung, Hochschule Mannheim
renan_veliz@yahoo.com

Abstract: Es wird ein aufwandsgünstiges Verfahren vorgestellt, welches das Schreien eines Babys in Abgrenzung zu anderen akustischen Signalen detektiert. Der betrachtete Anwendungskontext ist die Situation eines im Schlaf beunruhigten Babys, das zu weinen oder zu schreien beginnt. Die erfolgreiche Detektion kann zur Wiedergabe von beruhigender Musik oder zum Benachrichtigen der Eltern genutzt werden. Der entwickelte Algorithmus basiert auf den Merkmalen Energieverlauf, Stimmhaftigkeit und Stimmgrundfrequenz. Zur Ermittlung der Merkmale wurden verschiedene gängige Verfahren wie LPC-, cepstrale sowie spektrale Korrelationsanalyse auf ihre Eignung in diesem Kontext untersucht. Die Klassifikation erfolgt mit einem speziellen Fensterungsverfahren, das abschnittsübergreifend auf Basis eines Gedächtnisses mit Hysterese und variablen Schwellen arbeitet. So kann mit einstellbarer Verzögerung eine Entscheidung über das Vorliegen eines Babyweinens bzw. Babyschreis getroffen werden. Erste Experimente zur Ermittlung der erzielbaren Erkennungsraten zeigten für LPC- und cepstrales Verfahren vergleichbare Raten (91-92%). Zur Echtzeitimplementierung wurde das LPC-Verfahren wegen geringerem Aufwand gewählt.

1 Einleitung

Ausgangspunkt für die hier beschriebenen algorithmischen Untersuchungen ist die Situation eines im Schlaf beunruhigten Babys, das zu weinen oder zu schreien beginnt. An der Hochschule Mannheim wurde eine Miniaturschaltung mit Mikrofon und Miniatur-Lautsprecher entwickelt, die bspw. in ein Kuscheltier integriert werden könnte. Sie soll diese Situation akustisch detektieren und mit einer Schlafmelodie oder mit aufgezeichneten beruhigenden Worten der Eltern reagieren. Überdies könnten ähnlich einem Babyphon die Eltern benachrichtigt werden.

Die Aktivierung in gängigen Babyphones erfolgt zumeist rein pegelgesteuert. Eine Spezifizierung auf die Stimmcharakteristik des Babys stellt in preisgünstiger Hardware die Ausnahme dar [4]. Auch Untersuchungen zur Signalcharakteristik von Babyschreien generell sind vergleichsweise selten und eher mit Fragen der medizinischen Diagnostik verknüpft, z.B. [1], [2], [3].

Ziel unseres Ansatzes ist es im gegebenen Anwendungskontext, eine spezifische Detektion von Babyweinen oder –schreien zu realisieren, die gegenüber anderen Schallquellen unempfindlich reagiert. Ein wichtiges Kriterium bei der Verfahrensentwicklung war überdies von vorne herein eine geringe Komplexität, da der Einsatz auf der obigen kostengünstigen, portablen Hardware gewünscht ist. So wurde in einem ersten Ansatz die Analyse der Merkmale Energieverlauf, Stimmhaftigkeit und Stimmgrundfrequenz gewählt. Im Folgenden wird zunächst das entwickelte Verfahren im Überblick vorgestellt. Es folgt die detaillierte Beschreibung der drei algorithmischen Alternativen zur Merkmalsschätzung sowie des Fensterungsverfahren. Abschließend werden dann die Alternativen auf Basis von Proben von Babyschreien sowie Erwachsenen- und Kinderstimmen evaluiert und verglichen sowie mögliche weitere Verbesserungen diskutiert.

2 Gesamtverfahren

Das entwickelte Verfahren gliedert sich gemäß Abb. 1 in vier Stufen.

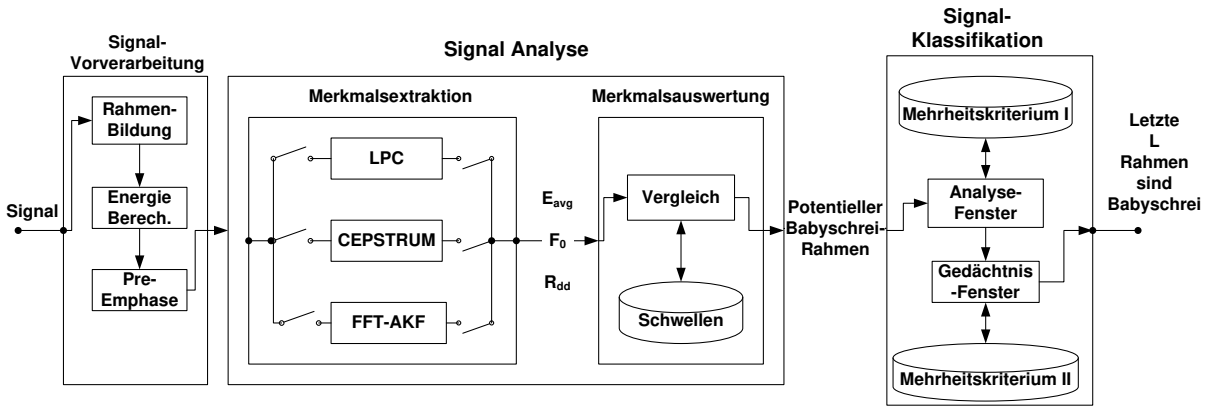

Abbildung 1 - Flußdiagramm des Verfahrens

Im ersten Schritt wird das mit 8 kHz abgetastete Signal normalisiert, Preemphase-gefiltert und in Rahmen von N=160-256 Abtastwerten gegliedert. Als erstes einfaches Klassifikationsmerkmal wird der mittlere Energiegehalt des Rahmens ermittelt. Ein Schwellwertvergleich schließt bereits hier energiearme Rahmen von der weiteren Analyse aus.

In der zweiten Stufe erfolgt die Schätzung der Merkmale Stimmhaftigkeit und Stimmgrundfrequenz. Im Rahmen dieser Arbeit wurden dazu drei alternative Algorithmen implementiert und untersucht, die in Kap. 3 ausführlicher dargestellt werden.

In der dritten Stufe werden die ermittelten Merkmale bewertet. Soll ein Rahmen für die abschließende vierte Klassifikationsstufe als potentieller Babyschrei-Rahmen in Betracht gezogen werden, so muss er drei Kriterien erfüllen, die wie folgt definiert wurden:

Kriterium 1: Ausreichender mittlerer Energiegehalt (E_{avg})

Im Unterschied zu leisen Umgebungsgeräuschen besitzen Babyschreie einen Mindestgehalt an Energie. Mit der Festlegung einer Schwelle für die mittlere Energie (E_{avg}) beispielsweise bei 70 dB erspart man sich die unnötige Analyse von Rahmen die höchstwahrscheinlich nicht stimmhaft sein werden.

Kriterium 2: Ausreichende Magnitude des Stimmhaftigkeits-Koeffizienten (R_{dd})

Babyschreie sind durch einen hohen Grad an Stimmhaftigkeit gekennzeichnet. Bei der Suche nach der Stimmgrundperiode ermitteln die drei in Kap. 3 betrachteten Algorithmen innerhalb eines vorgegebenen Suchintervalls das Maximum einer normierten Funktion f[lag]. Die Magnitude dieses Maximums dient hier als Maß für die Stimmhaftigkeit eines Rahmens in Form des Stimmhaftigkeits-Koeffizienten R_{dd} = f[maxlag].

Das Kriterium zur Stimmhaftigkeit kann durch eine der beiden folgenden Bedingungen erfüllt werden, wobei der Index j den aktuellen Rahmen und j-1 bzw. j-2 Vorgänger kennzeichnen, sowie R_{dd_low}:eine untere und R_{dd_up} eine obere Schwelle darstellen.

1. $R_{dd}[j] > R_{dd_low}$ && ($R_{dd}[j-1] > R_{dd_up}$ || $R_{dd}[j-2] > R_{dd_up}$)
2. $R_{dd}[j] > R_{dd_up}$ && ($R_{dd}[j-1] > R_{dd_low}$ || $R_{dd}[j-2] > R_{dd_low}$)

Kriterium 3: Passende Stimmgrundfrequenz F_0

Die Stimmgrundfrequenz für Babyschrei-Rahmen soll eine der folgenden Bedingung erfüllen:

1. $F_0[j] > F_{0_up}$
2. $F_0[j] > F_{0_low}$ && ($F_0[j-1] > F_{0_low}$) || ($F_0[j-2] > F_{0_low}$)

Die erste Bedingung ist erfüllt, wenn die ermittelte Stimmgrundfrequenz einen großen Abstand zum F_0 normaler Erwachsenenstimmen hat (z.B. 400 Hz). Die zweite Bedingung

erlaubt ein niedrigeres F_0, konditioniert dieses aber auf das Überschreiten von Schwellen in den beiden Vorgängerrahmen.

In der vierten Stufe des Verfahrens wird die Klassifikationsentscheidung mit Hilfe zweier rahmenübergreifender Fenster getroffen. Das entwickelte Vorgehen wird in Kap. 4 beschrieben.

3 Merkmalsextraktion

Zur Merkmalsextraktion wurden drei Alternativen untersucht.

3.1 LPC-Algorithmus

Diese Verfahrensvariante basiert auf einer LPC-Filterung (Ordnung 10) des Eingangssignals. Im Fehlersignal am Ausgang ist der Einfluss der Formanten weitgehend entfernt. Es entspricht somit der Anregung am Eingang des Vokaltraktes und enthält dementsprechend für stimmhafte Laute ein großes Maß an Periodizität nach Maßgabe der Stimmgrundfrequenz. Zu deren Bestimmung wird hier die Autokorrelation berechnet:

$$R(n) = \sum_{k=0}^{\frac{N}{2}-1} x[k]x[k+n] \quad mit \quad (\mathrm{N} = \text{Rahmengröße})$$

Das Maximum wird in einem auf die erwartete Grundfrequenz abgestimmten Intervall gesucht:

$10 < No < N/2$ und im Frequenzbereich $800\ Hz > \frac{Fs}{No} > F_{0_Low}$ mit F_{0_Low} z.B. 300 Hz

Hauptkriterium für das Entdecken des Maximums ist die Annahme, daß das zur Stimmgrundfrequenz gehörige Maximum notwendigerweise das erste ist, wenn man die Suche bei Lag 0 beginnt und in Richtung steigender Lags läuft. Meistens ist dieses Maximum auch dasjenige mit dem größten Betrag:

$$F_0 = \frac{Fs}{No} \quad \text{wobei} \quad No = arg\{\, max\{\, AKF[n]\,\}\,\}$$

In Ausnahmefällen wird auf diese Weise jedoch ein „falsches“ Maximum gefunden, obwohl die Suche nach der Stimmgrundfrequenz hier im LPC-Fehlersignal und nicht direkt im Sprachsignal durchgeführt wird, da es bei letzterem wahrscheinlicher ist Sub- oder Supra-Harmonische von F_0 zu detektieren. Die Gründe hierfür können in stark ausgeprägten ersten Formanten liegen. Ebenso kann es vorkommen, dass das Grundfrequenz-Maximum zu eng am ersten Formanten liegt und das Analysefilter dieses mitentfernt.

Als Maß für die Stimmhaftigkeit eines Rahmens wird hier der normierte AKF Koeffizient am gefundenen Maximum genutzt. Der Stimmhaftigkeits-Koeffizient für Kriterium 2 der Merkmalsauswertung wird dementsprechend zu R_{dd} = AKF[N_0] / AKF[0] berechnet. Die Robustheit dieses Maßes kann durch das Einbeziehen weiterer Parameter verbessert werden [5]: a) Abprüfen der Stimmhaftigkeit auf nebeneinander liegenden Rahmen, b) Nulldurchgangsrate im Rahmen, c) Energie des Prädiktionsfehlersignals. In Kriterium 2 wurde im Sinne der aufwandsarmen Implementierung nur die erste Erweiterung einbezogen.

3.2 CEPSTRUM-Algorithmus

Gemäß **[5]** wurde das Cepstrum nicht mit der üblichen Formel berechnet, sondern mit:

$$Ceps = IFFT\left\{[FFT\,(x[n])]^{\frac{1}{4}}\right\}$$

Nach einer FFT wendet man die 4te-Wurzelfunktion anstatt des Logarithmus an. Dies hat den Effekt, die Magnituden mit Betrag zwischen 0 und 1 näherungsweise linear abzubilden, während bei der Logarithmusfunktion solche Werte gegen -∞ streben. Dies würde negative Spitzen erzeugen, die das Ergebnis der folgende IFFT Operation beeinträchtigen. Im

Cepstrum ist die Übertragungsfunktion der Formanten deutlich von der Stimmgrundfrequenz getrennt, welche an dem vom Ursprung relativ weit entfernten Maximum erkennbar ist. Die Maximumlage ist mit einem Lag an einer AKF vergleichbar. Die Grundfrequenz errechnet sich entsprechend:

$$F_0 = \frac{F_S}{N_0} \quad \text{wobei} \quad N_0 = arg\{\, max\{\, CEPS[n] \,\}\,\}$$

Die Stimmhaftigkeit des bearbeiteten Rahmens wird durch die Magnitude des bei der Quefrenz N_0 im Cepstrum gefundenen Maximums bestimmt. Dieser Koeffizient wird nicht ausdrücklich normiert wie derjenige der AKF des LPC-Fehlers, da eine indirekte Normierung an der FFT des Eingangssignals stattfindet. Auch beim Cepstrum wirkt die Berücksichtigung der Stimmhaftigkeit nebeneinander liegender Rahmen verbessernd auf die Stimmhaftigkeitsentscheidung des gegenwärtigen Rahmens.

3.3 FFT-Algorithmus

Der FFT-AKF-Algorithmus sucht die Stimmgrundfrequenz direkt im FFT-Spektrum. Schwierigkeiten ergeben sich aber durch den bei stimmhafter Sprache häufig sehr starken ersten Formanten, mit welchem das erwartete Maximum bei der Grundfrequenz eines Babyschreis verwechselt werden kann, da die stimmhafte Energie auf die Harmonischen verteilt ist und so das spektrale Maximum der Stimmgrundfrequenz häufig energieärmer als der Formant ist. Die Grundfrequenz spiegelt sich aber auch im konstanten Abstand zwischen den harmonischen Vielfachen wider. Ein Ansatz, die Grundfrequenz aus dieser fast unveränderten Distanz zwischen Maxima im Spektrum zu gewinnen, ist das Harmonic Product Spectrum (HPS) [6].

Der FFT-AKF-Algorithmus versucht diesen Abstand mit einer Autokorrelation des FFT-Spektrums und anschließender Maxima-Suche zu gewinnen. Die Bestimmung der Stimmgrundfrequenz geschieht somit wie folgt:

$$F_0 = \frac{F_S}{N_0} \quad \text{wobei} \quad N_0 = \frac{NFFT}{N_0'} \quad \text{und} \quad N_0' = \arg\{\, \max\{\, AKF(FFT[n]) \,\}\,\}$$

Die Suche erfolgt im Intervall von Lag 10 bis Lag NFFT/2, wobei NFFT der Anzahl der Stützpunkte der FFT (z.B. 256) entspricht. N_0‘ ist der „Lag“ in der AKF des Spektrums („Frequenzlag“), der ungleich Null ist und den größten Wert besitzt. Wie beim LPC-Algorithmus wird auch hier das Verhältnis R_{dd} = AKF[N_0] / AKF[0] als Maß für die Stimmhaftigkeit des Rahmens herangezogen.

4 Signal-Klassifikation

Die abschließende Klassifikation des Signals wird durch die fortlaufende Analyse zweier Fensterverschiebungen und ihrer sich mitverändernden Gedächtnisse realisiert. Das Verfahren nutzt die durch die Merkmalsbewertung ermittelte Charakterisierung potentieller Babyschrei-Rahmen, um die Entscheidung zu treffen, ob es sich bei dem aktuellen Signalabschnitt um einen Babyschrei handelt.

Dazu werden zwei fortlaufende Fenster, Analysefenster und Gedächtnis-Fenster, gemäß Abb. 2 definiert. Die Abarbeitung der Prozesse Analyse-Fensterung und Gedächtnis-Fensterung erfolgt seriell.

Die im Flussdiagramm mit Rauten abgebildeten Abfragen beziehen sich auf einstellbare Schwellen, die die Verfahrensdauer bestimmen. Bei der Analyse-Fensterung ist die Berechnung des Verhältnisses zwischen der im Fenster enthaltenen Anzahl potentieller Babyschrei-Rahmen und stimmhafter Rahmen das Maß für die Markierung der Stelle bzw. des Rahmens, an dem dieses Verhältnis (Verhältnis_1) signifikant geworden ist, nämlich beim Überschreiten der Schwelle für das Verhältnis_1 (Analyse-Schwelle). Ein Zähler im Gedächtnis-Fenster registriert fortlaufend, wie viele Rahmen im Fenster ein Babyschrei-

Ereignis aufweisen. Die Aufzählung dieser Rahmen in Bezug auf die Länge des Gedächtnis-Fensters ergibt ein neues Verhältnis (Verhältnis_2), das ausdrückt, zu welchem Prozentanteil, die Rahmen im Gedächtnis-Fenster einem Babyschreisignal angehören.

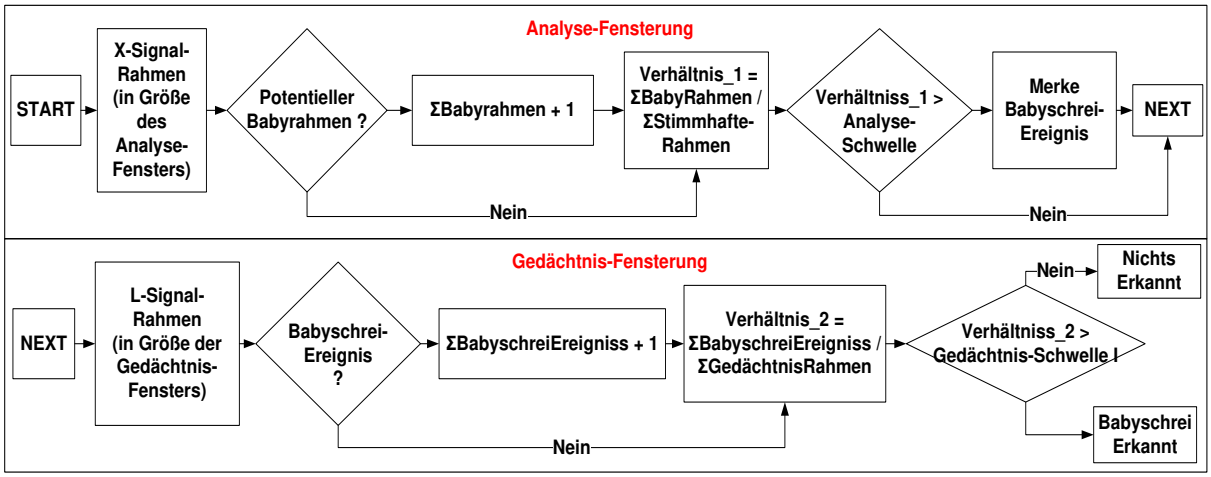

Abbildung 2 - Flussdiagramm des Algorithmus zur kontinuierlichen Zeitfensterung. Oberer Strang: Analyse-Fensterung; Unterer Strang: Gedächtnis-Fensterung.

Das Ergebnis der Analyse im Gedächtnis-Fenster und damit auch des gesamten Babyschreierkennungs-Algorithmus, ist die Entscheidung, ob das Signal ein Babyschrei ist oder nicht. Diese Entscheidung bezieht sich auf den gegenwärtigen Rahmen samt einer der Länge des Gedächtnis-Fensters entsprechenden Anzahl von Vorgängerrahmen. Abschließend wird eine Hysterese realisiert, indem erst eine Senkung des Verhältnis_2 unter eine niedrigere Hysterese-Schwelle einen Wechsel zur Entscheidung „Nicht-Babyschrei“ bewirkt.

Für die nachfolgenden Experimenten wurden das Analyse-Fenster auf 20 Rahmen (640 ms) und das Gedächtnis-Fenster auf 100 Rahmen (3,2 Sekunden) eingestellt. Justierungen dieser Größen sind je nach Anwendungssituation möglich und nötig zur Gesamtoptimierung.

5 Experimente

5.1 Set-up

In einem ersten Test wurden insgesamt 15 Audiodateien (WAV-Format) mit einer Gesamtlänge von 402 Sekunden benutzt. 7 Dateien bestehen aus reinen Babyschreien (201 Sekunden). Weitere 2 Dateien sind Sprachsignale von sprechfähigen Kindern (1 Junge und 1 Mädchen). Die restlichen 6 Dateien sind Sprachsignale von Erwachsenen. Sie unterteilen sich in männliche und weibliche Stimmen von bekannten Politikern: Barack Obama, Arnold Schwarzenegger und Tony Blair, sowie Hillary Clinton, Rania of Jordan und Sarah Palin. Alle stammen aus dem Internet. Umfangreichere Untersuchungen sind in Vorbereitung. Dazu werden Aufnahmen von „echten“ Babyschreien in einer Kinderklinik gemacht.

Die vorliegenden Stimmaufnahmen wurden zur Erstellung von Referenzen visuell an Hand von Spektrogrammen mit Angabe von Energieverlauf und Stimmgrundfrequenz-Kontouren evaluiert, auf ihren Klang beurteilt und manuell annotiert. Die Annotation ordnet jeden Rahmen einer der 5 Kategorien Rauschen, Babyschrei, Männerstimme, Frauenstimme oder Kinderstimme zu. Zur Analyse im Sinne einer binären Klassifikation, wie sie durch das vorgestellte Verfahren realisiert wird, können alle Kategorien außer der des Babyschreis, zur Kategorie „Other“ zusammengefasst werden. Die Auswertung mit mehreren Kategorien ermöglicht die Analyse darüber, mit welchen anderen Schallquellen die verschiedenen Algorithmus-Varianten einen Babyschrei-Rahmen verwechseln.

Vor der eigentlichen Testdurchführung wurden umfangreiche Untersuchungen durchgeführt, um die verschiedenen Schwellen im Gesamtverfahren zu justieren.

5.2 Ergebnisse und Bewertung

Im Folgenden werden die ermittelten Erkennungsraten für die drei Algorithmus-Varianten dargestellt. Dazu wurde die rahmenweise binäre Klassifikationsentscheidung (Baby-Schrei oder Nicht) der Algorithmen mit den Annotationen in den Referenzdateien verglichen. Für die getesteten Algorithmen im Einzeln wurden 12759 Rahmen @ 256 Abtastwerten verarbeitet.

TRUE \ RATIO(%)	RECOGNIZED AS: BABY	RECOGNIZED AS: OTHER
BABY	91	9
NOISE	39	61
MAN	0	100
WOMAN	0	100
KID	71	29

TRUE \ RATIO(%)	RECOGNIZED AS: BABY	RECOGNIZED AS: OTHER
BABY	91	9
OTHER	25	75

Abbildung 3 - Gütebeschreibung des binären Klassifikator für den LPC-Algorithmus

TRUE \ RATIO(%)	RECOGNIZED AS: BABY	RECOGNIZED AS: OTHER
BABY	92	8
NOISE	39	61
MAN	0	100
WOMAN	0	100
KID	62	38

TRUE \ RATIO(%)	RECOGNIZED AS: BABY	RECOGNIZED AS: OTHER
BABY	92	8
OTHER	24	76

Abbildung 4 - Gütebeschreibung des binären Klassifikators für den CEPSTRUM-Algorithmus

TRUE \ RATIO(%)	RECOGNIZED AS: BABY	RECOGNIZED AS: NOISE
BABY	54	46
NOISE	24	76
MAN	0	100
WOMAN	0	100
KID	53	47

TRUE \ RATIO(%)	RECOGNIZED AS: BABY	RECOGNIZED AS: NOISE
BABY	54	46
NOISE	17	83

Abbildung 5 - Gütebeschreibung des binären Klassifikators für den FFT-AKF-Algorithmus

Der LPC-Algorithmus hat eine hohe Sensitivität (91%), aber eine vergleichsweise niedrigere Spezifität (75%). Der Ergebnisse des Cepstrum-Algorithmus liegen mit einer Sensitivität von 92% und einer Spezifität von 76% leicht besser. Der Unterschied ist jedoch kaum signifikant. Der FFT-AKF-Algorithmus hat insgesamt eine niedrigere Sensitivität (54%) aber eine bessere

Spezifität (83%) als die beiden ersten Algorithmen. Für den praktischen Einsatz ist diese Sensitivität für Babyschreie nicht ausreichend. Man müsste starke Verbesserungen realisieren.

Für die Umsetzung in eine günstige Hardware ist neben der Erkennungssicherheit auch der algorithmische Aufwand zu bewerten. Beim Cepstrum ist hier ein Logarithmus bzw. eine vierte Wurzel zu errechnen. Viele DSPs arbeiten optimal mit dem Fractional 1.15 Datenformat. Die Cepstrum-Berechnung könnte hier u.U. problematisch werden. Der LPC-Algorithmus dagegen ist unkritisch in der Implementierung und bietet zudem Möglichkeiten zur einfachen Bestimmung weiterer Merkmale, bspw. der Residualfehlerenergie. Er wurde daher ausgewählt, in einer Hardware mit der Programmiersprache C eingesetzt zu werden.

5.3 Möglichkeiten zur Verbesserung

Betrachtet man die Art der auftretenden Verwechslungen stellt man fest, dass die Abgrenzung gegen Männerstimmen einwandfrei ist. Dies ist auf Grund der klar niedrigeren Stimmgrundfrequenz nicht verwunderlich. Unterscheidungsprobleme können jedoch bei Stimmen von Kindern und seltener auch bei Frauenstimmen auftreten. Auch bei Musik ist nicht auszuschließen, daß aufgrund von puren Sinustönen der Algorithmus falsch positiv reagiert. Verbesserungen sind mit Frauen- und Kinderstimmen zu machen, da diese eine hohe Stimmgrundfrequenz besitzen, was zum Teil zu einer Fehlschätzung der F_0 führt. Hierzu können die definierten Schwellen optimiert werden, jedoch liegt ein prinzipielles Problem darin, dass sich die typischen F_0-Intervalle von Kinderstimmen (teilweise auch Frauenstimmen) und Babyschreien überlappen.

Verbesserungen der Algorithmen können generell an den Stellen gemacht werden, wo eine Maximasuche stattfindet, da hier immer wieder Fehler auftreten. Bspw. könnte das LPC-Fehlersignal Bandpass-gefiltert werden, um bereits das Signal gezielt auf das Frequenzintervall der Babystimmgrundfrequenz zu begrenzen. Gegenüber dem bereits durchgeführten Begrenzen des Suchintervalls hätte dies den Vorteil, dass Formanteffekte weiter unterdrückt werden könnten.

Die Messung des Energiegehalts des LPC-Fehlers könnte auch dazu beitragen einen Sinuston besser von einem Babyschrei zu unterscheiden, da das LPC-Filter diesen fast komplett vorhersagen kann. Selbst wenn im Prädiktionsfehler eine typische Sinusoid-Periodizität übrig bleibt und die AKF auf eine Grundfrequenz hinweist, wäre die enthaltene Energie sehr gering.

Der FFT-AKF-Algorithmus versucht mit der AKF des Spektrums die Periodizität der harmonischen Komponenten zu erkennen. Das Maximum, das die Grundfrequenz repräsentiert, ist jedoch häufig nicht deutlich genug ausgeprägt, da die Energie auch in den höheren Harmonischen verteilt ist. Dies führt dann zu einer falschen Stimmgrundfrequenzbestimmung. Ein Ansatz zur Verbesserung wäre die Periodizität an der FFT der AKF zu erkennen. Dies bedeutet, die Anwendung einer weiteren FFT auf die AKF mit entsprechend erhöhtem Rechenaufwand. Einfachere Alternativen (z.B. AMDF) mit der Effektivität einer FFT wären möglich. Ähnliches wäre auch im Cepstrum anwendbar, um die rahmonischen Quefrenzen zu berücksichtigen.

6 Zusammenfassung und Ausblick

Im Ergebnis der hier dargestellten Untersuchungen wurde der LPC-Algorithmus für die Implementierung in einer echtzeitfähigen Miniaturschaltung ausgewählt. Diese Entscheidung wurde aufgrund der guten Erkennungsrate und der geringen Komplexität des Verfahrens getroffen. Durch die Vielzahl konfigurierbarer Schwellwerte lässt sich das Verfahren auf die akustischen Bedingungen der jeweiligen Anwendungssituation anpassen.

Eine Reihe von Ansätzen zur Verbesserung des Verfahrens wurde in Kap. 5.3 skizziert. Das prinzipielle Problem, dass sich die typischen Stimmgrundfrequenz-Intervalle von Babyschreien mit denjenigen von Kinderstimmen (teilweise auch Frauenstimmen)

überlappen, lässt sich jedoch durch alleinige Nutzung der drei vorgestellten Merkmale nicht lösen. Man bräuchte einen weiteren Parameter, der Babyschreie besonders charakterisiert bzw. im Babyschreisignal einzigartig ist.

Erste weitergehende Untersuchungen zeigen Verbesserungspotential durch das Einbeziehen des Formantenverlaufs auf. Das Verhalten einzelner Formanten (z.B. des Ersten) über der Zeit, kann als weiteres wesentliches Kriterium dienen, um eine zuverlässige Aussage über das Vorliegen eines Babyschreis zu treffen. Grundlage dafür ist, dass das Baby beim Schreien den Vokaltrakt weniger stark über der Zeit variiert, als ein Kind oder Erwachsener dies bei der Artikulation während des Sprechens tut. Dieses Verhalten lässt sich beobachten, wenn man die Zeitverläufe der ersten Formanten während eines Babyschreis mit denen von Sprache vergleicht. Die Formantenspur von Sprache weist eine deutlich größere Varianz auf. Die fortlaufende Messung dieser Schwankungen kann ein Maß liefern, das als weiteres Merkmal den Unterschied zwischen Babyschrei und Kinder-Sprache hervorhebt. Jedoch ist diese Analyse über einer Mindest-Zeitspanne zu betreiben und für eine Implementierung in einer kostengünstigen Hardware mit beträchtlichem Aufwand verbunden.

Literatur

[1] Lederman, Dror: Automatic Classification of Infant's Cry. Master Thesis. Ben-Gurion University of the Negev Faculty of Engineering Sciences, Department of Electrical and Computer Engineering. October 2002.

[2] Reyes-Galaviz, Orion F. & Reyes-García, Carlos.A.: A System for the Processing of Infant Cry to Recognize Pathologies in Recently Born Babies with Neural Networks, Instituto Tecnólgico de Apizaco, Tlaxcala, México. SPECOM'2004: 9th Conference Speech and Computer St. Petersburg, Russia September 20-22, 2004

[3] Möller, Sebastian & Schonweiler, Rainer: "Analysis of infant cries for the early detection of hearing impairment", In EUROSPEECH-1997, 1997

[4] Hsieh, Chau-Kai: US Patent 5668780 - Baby cry recognizer, September 16, 1997. http://www.patentstorm.us/patents/5668780/claims.html

[5] Benesty, M., Mohan Sondhi, Yiteng Huang. Springer Handbook of Speech Processing 2008

[6] Middleton, Gareth: Pitch Detection Algorithms, 2003. http://cnx.org/