



Doctoral Thesis

Algorithms for sound classification in hearing instruments

Author(s):

Büchler, Michael Christoph

Publication Date:

2002

Permanent Link:

<https://doi.org/10.3929/ethz-a-004281408> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH no 14498

Algorithms for Sound Classification in Hearing Instruments

A dissertation submitted to the

SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZURICH

for the degree of

Doctor of Technical Sciences

presented by

Michael Christoph Büchler

dipl. El. Ing. ETH

born April 26, 1967

citizen of Brugg/AG

accepted on the recommendation of

Prof. Dr. Peter Niederer, examiner

PD Dr. Norbert Dillier, co-examiner

Dr. Stefan Launer, co-examiner

2002

Abstract

Hearing instrument users often prefer different instrument-settings in different acoustic environments. Thus, modern hearing instruments allow the user to select between several hearing programs for different situations, to change the frequency response and compression parameters, or to activate a directional microphone, noise reduction or feedback suppression. However, the user has the bothersome task of recognizing the acoustic environment and then switching to the program that best fits this situation, using a switch on the hearing instrument or a remote control. Automatic sensing of the current acoustic situation and automatic switching to the best fitting program would therefore greatly improve the utility of today's hearing instruments.

The above assumption was confirmed by practical experiences. In a study with hearing impaired subjects, the usefulness and acceptance of an automatic program selection mode in the hearing instrument was investigated from the point of view of the user. It was shown that the automatic switching mode of the test instrument was deemed useful by a majority of test subjects, even if its performance was not perfect. These results were a strong motivation for the research described in the present work.

In this thesis, an automatic sound classification system for application in hearing instruments is developed. Our goal was on the one hand to develop a robust classification algorithm for at least the four classes 'speech', 'speech in noise', 'noise', and 'music', on the other hand to collect fundamental knowledge as a basis for a more detailed classification. The refined classes may for example be different noise types and music styles, or clean and reverberated speech.

So far, existing sound classification algorithms designed for hearing instruments have particularly been able to separate speech signals from other signals. Musical sounds however could not be recognized, and it was only partly possible to separate noise from speech in noise. Other algorithms, designed primarily for multimedia applications, allowed the recognition of sounds on a more specific level, such as certain distinct alarm signals or certain kinds of music. Thus, for the recognition of the four more general sound classes mentioned above, a new approach had to be developed. Generally, a sound classification system consists of the extraction of appropriate features from the signal, followed by a pattern classifier and an optional post processing step. This architecture was also chosen for the system described in this thesis.

In order to study how the human auditory system classifies sound, the mechanisms of Auditory Scene Analysis were investigated. The extraction of auditory features was shown to be an important step in the process of sound segmentation performed by the human auditory

system. Thus, one of the main goals in this thesis was to find appropriate features. A number of adequate auditory features have been modeled, including amplitude modulations, harmonicity, spectral profile, amplitude onsets, and rhythm.

These auditory features were evaluated together with different pattern classifiers. Considering the application in hearing instruments, where computing time and memory are limited, simple classifiers (rule-based and minimum-distance classifiers) have been compared with more complex ones (Bayes classifier, neural network, hidden Markov model, and a multistage approach). A hit rate of about 80 % was achieved with the simpler classifiers, which could be increased up to some 90 % when a more complex classifier was used. However, both the computing time and memory requirements are about four times larger with the more complex than with the simpler approaches.

The best classification system contains two stages: The first stage consists of a first feature set and a hidden Markov model. In the second stage, comprising a second feature set and a rule-based approach, the output of the hidden Markov model is verified and the classification is corrected if necessary. This approach worked well for most sounds within the four classes, resulting in a hit rate of 91 %. There are a number of sounds in each of the four classes that were recognized very robustly: Clean and slightly reverberated speech, speech in noise with moderate SNR, traffic and social noise, and classical music, single instruments and singing. However, some sounds were problematic and were mostly misclassified: Speech in noise with very low or very high SNR was classified as 'noise' or 'speech', respectively; compressed and strongly reverberated speech, a few tonal and fluctuating noises, and compressed pop music were all classified as 'speech in noise'.

Thus, for a more detailed sound classification in hearing instruments, further research is required. However, some fundamental limitations are already evident: the hearing instrument will not always be able to recognize whether a sound will be regarded as a desirable signal or as noise by the user.

Zusammenfassung

Benutzer von Hörgeräten bevorzugen abhängig von der akustischen Umgebung oft unterschiedliche Geräteeinstellungen. Darum ermöglichen es moderne Hörgeräte auch, mehrere Programme für verschiedene akustische Situationen zu wählen, um den Frequenzgang und Kompressionsparameter zu ändern, oder Richtmikrofon, Störgeräusch-Reduktion oder Feedback-Unterdrückung zu aktivieren. Allerdings müssen die Hörgeräteträger bei herkömmlichen Mehr-Programm-Hörgeräten selbständig die akustische Hörsituation beurteilen und entscheiden, welches Programm für diese Situation optimal ist, um dann per Schalter am Hörgerät oder über eine Fernbedienung in das entsprechende Programm umzuschalten. Eine automatische Erkennung der aktuellen akustischen Situation und automatisches Umschalten in das geeignetste Programm würden deshalb den Komfort von solchen modernen Hörgeräten verbessern.

Diese Annahme wurde durch praktische Erfahrungen bestätigt. In einer Studie mit Schwerhörigen wurden die Nützlichkeit und Akzeptanz einer automatischen Programmwahl in Hörgeräten aus der Sicht der Benutzer untersucht. Es konnte gezeigt werden, dass die Automatik des Testinstruments von einer Mehrheit der Versuchspersonen als nützlich beurteilt wurde, auch wenn die Programmwahl nicht immer korrekt war. Diese Resultate waren eine grosse Motivation für die in dieser Dissertation beschriebene Forschung.

In der vorliegenden Arbeit wird ein System zur Klangklassifizierung für die Anwendung in Hörgeräten entwickelt. Das Ziel war einerseits, einen robusten Klassifizierungs-Algorithmus für mindestens die vier Geräuschklassen 'Sprache', 'Sprache im Störgeräusch', 'Störgeräusch' und 'Musik' zu entwickeln, andererseits grundlegende Erfahrungen als Basis für eine detailliertere Klassifizierung zu sammeln. Eine mögliche feinere Unterteilung der genannten Klassen sind zum Beispiel verschiedene Typen von Störgeräuschen, verschiedene Musikstile oder reine und verhallte Sprache.

Bisherige Ansätze zur Klang-Klassifizierung in Hörgeräten waren insbesondere dazu fähig, Sprache von anderen Signalen zu trennen. Musik jedoch konnte nicht erkannt werden, und es war nur beschränkt möglich, Störgeräusch von Sprache im Störgeräusch zu trennen. Weitere Geräuschklassifizierungs-Algorithmen wurden vor allem für Multimedia-Applikationen entwickelt; sie erlauben es, spezifischere Klänge zu erkennen, wie zum Beispiel einzelne Alarmsignale oder gewisse Musikstile. Für die automatische Erkennung der vier oben erwähnten Geräuschklassen musste daher ein neuer Ansatz entwickelt werden. Im allgemeinen besteht ein System zur Geräuschklassifizierung aus der Extraktion von geeigneten Merkmalen aus dem Signal, gefolgt von einem Klassifizierer und einem optionalen Nachverarbeitungsblock. Diese Architektur wurde auch für das in der vorliegenden Arbeit entwickelte System verwendet.

Um darüber Kenntnis zu erlangen, wie das menschliche auditorische System Klangklassifizierung vornimmt, wurden die Mechanismen der auditorischen Szenenanalyse studiert. Es zeigte sich, dass die Extraktion von auditorischen Merkmalen eine wichtige Stufe im Prozess der Geräuscherkennung im auditorischen System ist. Daher war eines der Hauptziele der vorliegenden Arbeit die Suche nach geeigneten Merkmalen. Eine Reihe von zweckmässigen auditorischen Merkmalen wurde modelliert. Dazu gehören Amplitudenmodulationen, Harmonizität, spektrales Profil, Amplituden-Onsets und Rhythmus.

Diese auditorischen Merkmale wurden zusammen mit verschiedenen Klassifizierern evaluiert. Im Hinblick auf eine Anwendung in Hörgeräten, wo Rechenzeit und Speicherplatz beschränkt sind, wurden einfachere Klassifizierer (regelbasierter und Minimum-Distance Klassifizierer) mit komplexeren verglichen (Bayes Klassifizierer, neuronales Netz, hidden Markov Modell, und ein mehrstufiger Ansatz). Mit den einfacheren Ansätzen konnte eine Trefferquote von etwa 80 % erreicht werden, die auf bis zu 90 % erhöht werden konnte, wenn ein komplexerer Klassifizierer gewählt wurde. Sowohl der Bedarf an Rechenzeit als auch an Speicherplatz ist jedoch für die komplexeren Ansätze etwa vier mal höher als für die einfachen.

Das beste System umfasst zwei Stufen zur Klassifizierung: Die erste Stufe besteht aus einem ersten Merkmalsset und einem hidden Markov Modell. In der zweiten Stufe, bestehend aus einem zweiten Merkmalsset und einem regelbasierten Klassifizierer, wird das Resultat des hidden Markov Modells verifiziert und die Klassifizierung wenn nötig korrigiert. Dieses System erzielte eine Trefferquote von 91 %. Einige Klänge in jeder der vier Klassen wurden damit sehr robust erkannt. Dies sind reine und leicht verhallte Sprache, Sprache im Störgeräusch mit moderatem SNR, Verkehrs- und Partygeräusch, klassische Musik, einzelne Instrumente und Gesang. Einige Klänge waren jedoch problematisch und wurden meist falsch klassifiziert: Sprache im Störgeräusch mit ziemlich kleinem oder grossem SNR wurden als 'Störgeräusch' beziehungsweise 'Sprache' klassifiziert, komprimierte und stark verhallte Sprache, einige tonale und fluktuierende Störgeräusche und komprimierte Popmusik wurden alle als 'Sprache im Störgeräusch' betrachtet.

Für eine detailliertere Geräuschklassifizierung in Hörgeräteanwendungen ist somit weiterer Forschungsbedarf vorhanden. Dabei sind jedoch einige grundsätzliche Grenzen vorgezeichnet; das Hörgerät wird nicht immer erkennen können, ob der Benutzer ein Geräusch als Nutzsignal oder als Störsignal empfindet.