

Digitale Medien (DigMed)

Voice Activity Detection und Comfort Noise Generation

Studienarbeit von Samuel Huber und Michael Nipkow

Key Words: Studienarbeit, Voice Activity Detection (VAD), Comfort Noise Generation (CNG), VAD, CNG, Voice over IP (VoIP), G.723.1



Ausgabe Datum: 08.02.2005
Dokument Status: Freigegeben
Autoren: Samuel Huber, Michael Nipkow
Betreuer: Andreas Ehrensperger
Schule: HSR Hochschule für Technik Rapperswil

Änderungsgeschichte

Rev	Datum	Autor	Bemerkungen
1.3	08.02.2005	Michael Nipkow	- Überarbeitung und Korrektur
1.2	07.02.2005	Samuel Huber	- Dokumentation ergänzt
1.1	30.01.2005	Michael Nipkow	- Wiederaufnahme der Dokumentation
1	09.11.2004	Michael Nipkow	- Neuerstellung

2. Abstract

Im Rahmen dieser Studienarbeit ging es darum, ein System zu entwickeln, welches Sprache von Nicht-Sprache unterscheiden kann. Eine so genannte Voice Activity Detection – oder kurz VAD – kann heute überall dort eingesetzt werden, wo Sprache digital übertragen wird. Eine Anwendung, wo bereits Voice Activity Detection eingesetzt wird, ist das Mobiltelefon. Eine weitere Anwendung, die stark im Aufkommen ist, ist das so genannte Internet Telefon – auch IP-Telefon (IP für Internet Protocol) genannt. Oft wird auch nur die Bezeichnung Voice over IP verwendet, was für die selbe Technologie steht.

Um den Datenfluss zu reduzieren, werden in den Gesprächspausen (fast) keine Daten übertragen. Dass man den Datenfluss nicht ganz stoppen kann, rührt daher, dass Protokollheader, Steuerdaten und Zusatzinformationen trotzdem benötigt werden.

Ein Teil dieser Zusatzinformationen können zum Beispiel Informationen über das Hintergrundrauschen des Senders sein. Diese Informationen werden in den Gesprächspausen beim Sender erfasst und zum Empfänger übermittelt. Es wird nicht das eigentliche Rauschen als Sprache/Sound übertragen, sondern lediglich Parameter, welche dieses beschreiben. Mit diesen Parametern kann beim Empfänger in den Gesprächspausen ein „künstliches“ Rauschen generiert werden. Diese „Rauschgenerierung“ wird Comfort Noise Generation – oder kurz CNG – genannt, welche ebenfalls zum Umfang dieser Studienarbeit gehörte. Das „künstliche“ Rauschen soll verhindern, dass die Übergänge zwischen Sprache und Nicht-Sprache hörbar sind.

Ziel war es, einen Algorithmus für die VAD und die CNG zu entwickeln, diesen zu simulieren und schlussendlich auf dem DSP zu implementieren. Es ist uns zwar gelungen, einen lauffähigen Algorithmus zu entwickeln und zu simulieren, doch die Implementation auf dem DSP konnten wir aus mehreren Gründen leider nicht mehr realisieren.