

Richtungshören

Ein Richtungshörsystem für mobile Roboter in echoarmer Umgebung

Das Hörsystem beim Menschen

Das menschliche Hörsystem ist verhältnismäßig komplex aufgebaut, wenn man es mit anderen Sinnessystemen des Menschen vergleicht. Es ist so genau, dass man allein durch das Eingießen von Kaffee in eine Tasse unterscheiden kann, ob er heiß oder kalt ist. Außerdem dient es als primäres Alarm- und Lokalisationssystem. Sogar bei vielen verschiedenen Schallquellen ist es in der Lage, einzelne auszuweisen.

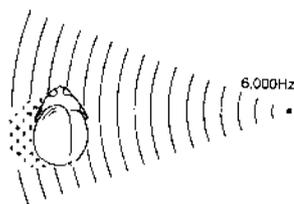
Verarbeitung:

Der Schall wird zuerst durch Außen- und Mittelohr geleitet. Anschließend werden die Signale im Innenohr in elektrische Impulse transformiert. Schließlich werden die Impulse vom Gehirn aufgenommen und neuronal verarbeitet.

Voraussetzungen:

Um überhaupt Richtungshören zu können, sind in jedem Fall 2 Ohren nötig. Hier werden unabhängig voneinander an zwei verschiedenen Orten Schallquellen aufgenommen. Mit Hilfe des Ohren-Abstandes können so Zeitunterschiede festgestellt werden, die der Schall benötigt um von einem Ohr zum anderen zu gelangen. Bei einer Schallgeschwindigkeit in Luft von ca. 340 m/s liegen die Zeitabstände zwischen den Ohren im μsec -Bereich. Diesen Zeitabstand nennt man auch ITD (Interaural Time Difference). Mit Hilfe des ITD kann der Mensch Schallquellen in einem Bereich von 180° lokalisieren, speziell links und rechts.

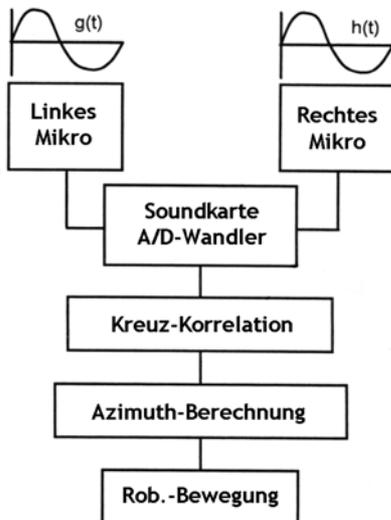
Um auch hinten, vorne, oben und unten (also schließlich alle 3 Dimensionen) zu lokalisieren ist noch ein 2. Prinzip nötig: die sogenannte Pegeldifferenz, auch ILD (Interaural Level Difference) genannt. Hierbei ist die Form der Ohren und des Kopfes entscheidend. Die Ohrmuschel beispielsweise bewirkt eine Verzerrung des Signals, abhängig davon aus welcher Richtung das Schallsignal kommt. Etwas anders wirkt sich die Kopfform aus.



Dieser stellt für die Schallwellen eine Art Hindernis dar, was dazu führt, dass die Schallquellen einen „Schatten“ werfen. Dies ist allerdings nur bei höheren Frequenzen (ab ca. 2-3 kHz) der Fall. Durch den Schatten-Effekt kommt es je nach Schall-Richtung zu Pegelunterschieden auf den Ohren, die sich auch zur Lokalisation nutzen lassen.

Für die Realisierung einer technischen Umsetzung eignet sich der ITD eher als der ILD, da zeitliche Abstände einfacher zu messen und erheblich einfacher auszuwerten sind als Pegeldifferenzen. Dafür muss man sich aber auf einen Lokalisationsbereich von 180° beschränken.

Eine mögliche technische Umsetzung



Eine Möglichkeit, das vorgestellte Prinzip technisch umzusetzen, ist auf der nebenstehenden Abbildung zu sehen. Hierbei nimmt man 2 Mikrophone mit einem fest definierten Abstand, misst damit die Schallquellen und versucht, den zeitlichen Abstand der eingetroffenen Schallquellen zu berechnen. Anhand des errechneten ITD kann man schließlich den Winkel der Schallquelle bestimmen (mit Beschränkung der 180°) und einen Roboter beispielsweise in diese Richtung drehen bzw. fahren lassen.

Der folgende Versuch wurde mit dem Roboter „PeopleBOT“ der Firma ActivMedia mit folgenden technischen Daten durchgeführt:

- ◆ Betriebssystem: Linux Red Hat
- ◆ K6-2-500 MHz CPU, 128 MB RAM
- ◆ 30 cm Mikrophon-Abstand
- ◆ Gewöhnliche Stereo-Soundkarte
- ◆ Abtastrate: 44,1 kHz
- ◆ Entspricht Genauigkeit von ca. 22 μsec

Der erste Schritt ist nun die Aufzeichnung der Schallquellen mit Hilfe der beiden Mikrophone. Diese zeichnet man für einen vorgegeben Zeitraum auf. Dieser muss mind. so lang sein wie die doppelte Zeit, die vergeht, wenn der Schall den Weg von dem einen zum anderen Mikrophon zurücklegt. Bevor die Daten aufgezeichnet werden können, müssen natürlich die analogen Signale von einer Soundkarte in digitale umgewandelt werden. Dazu genügt eine gewöhnliche Stereo-Karte mit einer Abtastfrequenz von 44,1 kHz, wie sie auch bei Audio-CDs üblich ist. Jede einzelne Abtastung hat so einen zeitlichen Abstand von $1/44100 \text{ sec} = \text{ca. } 22 \mu\text{sec}$. Eine sehr exakte Bestimmung des Winkels ist dadurch nicht gegeben, aber für Anwendungen bei denen keine hohe Präzision erforderlich ist, ist es vollkommen ausreichend. Man muss so mit Abweichungen von durchschnittlich $1,5^\circ$ rechnen.

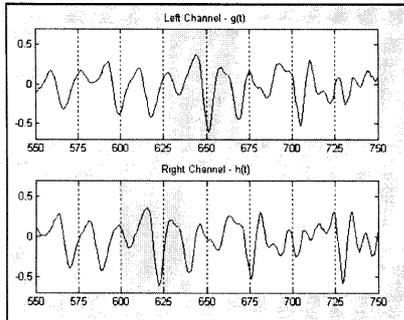
Die Kreuz-Korrelation

Mit Hilfe der Kreuz-Korrelation ist es möglich, etwa gleichverlaufende Kurven (z.B. in unserem Fall Schallsignale), die nur ineinander verschoben sind, den Abstand dieser Verschiebung zuverlässig zu erfassen.

Einfaches Verschieben und Vergleichen (so wie man es intuitiv machen würde) funktioniert leider nicht, da es leichte Differenzen der beiden Signale gibt; hervorgerufen durch Störungen aller Art, besonders bei der A/D-Wandlung in der Soundkarte. Also benötigt man eine Formel, die darauf anspricht, wenn eine Verschiebung gefunden wurde, bei der beide Kanäle am ehesten übereinstimmen.

Genau für diesen Zweck gibt es folgende Kreuz-Korrelations-Formel:

$$\text{Corr}(g, h)_j(t) \equiv \sum_{k=0}^{N-1} g_{j+k} h_k$$

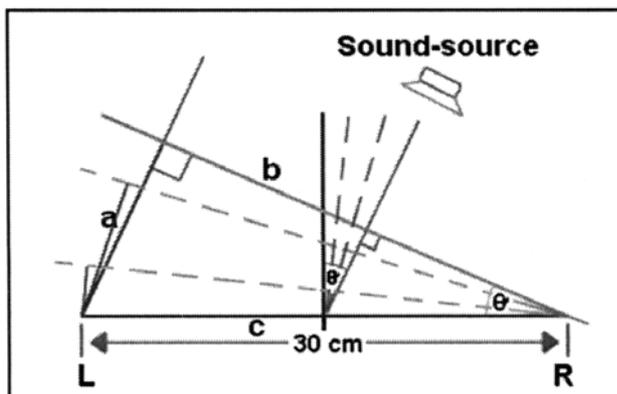


Als Eingabe erhält die Formel eine Liste von Signal-Werten der beiden Kanäle g und h . Diese Werte liegen im Bereich zwischen -1 und $+1$. Gesucht ist nun der Wert j , bei dem die Funktion das größte positive Ergebnis liefert. Der Wert j entspricht dabei gerade der Verschiebung des linken Kanals g . Geht man vorerst von zufällig verteilten Werten der beiden Kanäle aus, hat man zwei Listen von Werten zwischen -1 und $+1$. Nun werden alle Werte mit dem gleichen Index k (anders gesagt: die Werte, die zeitlich übereinanderliegen) der beiden Kanäle miteinander multipliziert und anschließend die Summe dieser Werte berechnet. Bei zufällig verteilten Werten ist somit

die Wahrscheinlichkeit für positive und negative Werte gleichverteilt, da bei Produkten von Zahlen mit gleichen Vorzeichen sich pos. Werte ergeben, bei Produkten von Zahlen mit verschiedenen Vorzeichen dagegen negative. Eine Summe von Werten die gleichverteilt positiv und negativ sind, ergibt statistisch gesehen genau Null. Dieser Fall entspricht bei unserem Schallmodell dem Fall, dass die beiden Kanäle sich sehr stark unterscheiden.

Wie verhält sich die Formel nun, wenn man den richtigen Wert j gefunden hat, und die beiden Kanäle annähernd übereinanderliegen. Geht man vom Ideal-Fall aus, dass die Werte der Kanäle g und h dann exakt gleich sind, erhält man folgendes: Der Fall von verschiedenen Vorzeichen bei den Produkten entfällt, da die Werte ja gleich sind und somit nur pos. Werte zustande kommen. Wenn nun die Summe dieser Werte berechnet wird, erhält man schließlich eine große positive Zahl, die von keinem anderen j übertroffen wird. Somit kann man tatsächlich sagen, dass das j , für den die Kreuz-Korrelation das größte pos. Ergebnis liefert, das gesuchte j ist, welches die zeitliche Verschiebung der beiden Kanäle angibt.

Die Azimuth Berechnung



Nun kann man mit dem Wert j , den man aus der Kreuz-Korrelation erhalten hat, den Azimuth-Winkel bestimmen. Die nebenstehende Abbildung bietet einen grafischen Überblick über das Problem. Mit Hilfe der Strahlensatz-Regeln kann man ein rechtwinkliges Dreieck konstruieren, das uns den Winkel relativ einfach berechnen lässt. Bekanntlich genügen in einem rechtwinkligen Dreieck 2 Seiten, um einen Winkel zu bestimmen. Die Seite c ist durch den Abstand der beiden Mikrophone gegeben. Des weiteren ist es möglich die Seite a folgendermaßen zu

berechnen: wir wissen, dass die Strecke a genau dem Weg entspricht, den der Schall noch zurücklegen muss, um das 2. Mikrophone zu erreichen. Da wir durch die Kreuz-Korrelation den zeitlichen Abstand errechnet haben, können wir in Kombination mit der Schallgeschwindigkeit die Strecke a bestimmen. Dies geschieht folgendermaßen:

$$\Delta = 1/44.1 \times 10^3 = 2.2676 \times 10^{-5} \text{ s}$$

$$t = \Delta \times \sigma$$

$$\begin{aligned} \text{length} &= t \times V_{\text{sound}} \\ &= (\Delta \times \sigma) \times V_{\text{sound}} \end{aligned}$$

$$\sin \Theta = \frac{a}{c} \therefore \Theta = \sin^{-1} \frac{a}{c}$$

$$\Theta = \sin^{-1} \frac{(\Delta \times \sigma) \times V_{\text{sound}}}{c}$$

Messergebnisse

In folgender Tabelle sieht man einige Ergebnisse des Versuchs. In der 1. Spalte stehen vorgegebene Winkel der Schallquelle. Die zugehörige theoretische Zeitdifferenz steht in der 2. Spalte. In der 3. Spalte findet man die theoretischen Werte, die die Kreuz-Korrelation berechnen sollte. Die eigentlich interessante Spalte ist die 4.. Hier sieht man die Werte, die die Kreuz-Korrelation nun tatsächlich berechnet hat. Die größte Mess-Abweichung beträgt 2 Einheiten bzw. ca. 5%. Die durchschnittliche Winkel-Abweichung des Roboters sieht man in der 5. Spalte. Mit max. 4° durchschnittliche Abweichung ist auch dieses Ergebnis akzeptabel. Hier kommen natürlich noch mechanische Ungenauigkeiten mit ins Spiel. Über alle Winkel gesehen, beträgt die durchschnittliche Abweichung sogar nur 1,5°.

Winkel	Zeitdifferenz	Kreuz-Korrelation (theoretisch)	Kreuz-Korrelation	Positions-Abw. nach Drehung
	[in μsec]	[j-Verschiebung]	[j-Verschiebung]	(Durchschnitt)
-90°	840	37	39	4°
-50°	599	26,4	27	2°
-30°	386	17	17	0°
0°	0	0	0	2°
10°	-136	-6	-5	2°
20°	-268	-11,8	-11	1°
45°	-554	-24,4	-26	2°
70°	-735	-32,4	-33	3°

Fazit

Wir haben ein System mit 2 Mikrofonen kennengelernt, das mit Hilfe von Schall-Laufzeitunterschieden eine Schallquelle im Bereich von 180° lokalisiert. Verwendet wurde dafür die Kreuz-Korrelation, die nach einigen Testversuchen offensichtlich zuverlässige Ergebnisse liefert und den Roboter der Genauigkeit der Ergebnisse entsprechend in die Richtung der Schallquelle drehen bzw. auch fahren lässt.

Quellen:

Jürgen Adamy, Kyriakos Voutsas, Volker Willert. Ein binaurales Richtungshörsystem für mobile Roboter in echoarmer Umgebung, 2003

John C. Murray, Harry Erwin, Stefan Wermter. Robotic Sound-Source Localization and Tracking using Interaural Time Difference and Cross-Correlation, 2004