Berücksichtigung der zeitlichen Wahrnehmung im Abbild natürlicher Geräusche

Eckard Blumschein

Otto von Guericke Universität Magdeburg
Institut für Elektronik, Signalverarbeitung und Kommunikationstechnik

Das Spektrogramm gilt als Abbild von Geräuschen. Nicht selten sind jedoch deutlich hörbare Merkmale wenig oder gar nicht ersichtlich. Dies gab Veranlassung nach Umfang und Ursachen entscheidender Defizite zu suchen sowie Korrekturmöglichkeiten zu erörtern. Es ging vor allem darum, in welchem Umfang und wie die zeitliche Wahrnehmung von natürlichen Geräuschen, Sprache und Musik in der Ebene u. a. mit Farben besser darstellbar gemacht werden kann.

1 Zeitliche Grenzen des Gehörs

Delphine und Menschen können Zeitabstände wahrnehmen, die weniger als eine bzw. vier Mikrosekunden betragen. Sie hören jedoch nur Frequenzen bis herauf zu 130 bzw. 20 kHz [1]. Dieses Paradoxon resultiert daraus, dass die Wechselkomponente des Rezeptorpotentials in den Haarzellen dem Stimulus zwar innerhalb von Bruchteilen einer Mikrosekunde biphasig folgt, Neurotransmitter jedoch jeweils nur bei Depolarisation zeitdiskret und gerichtet über einen synaptischen Spalt zu den afferenten Nervenfasern übertragen werden, wobei die Zahl der Haarzellen begrenzt ist. Letzteres erklärt auch warum ein Verlust von Hörzellen in Alter nicht die Frequenzgrenze sondern die Hörschwelle über einen weiten Bereich bis herab zu etwa 2 kHz verschiebt.

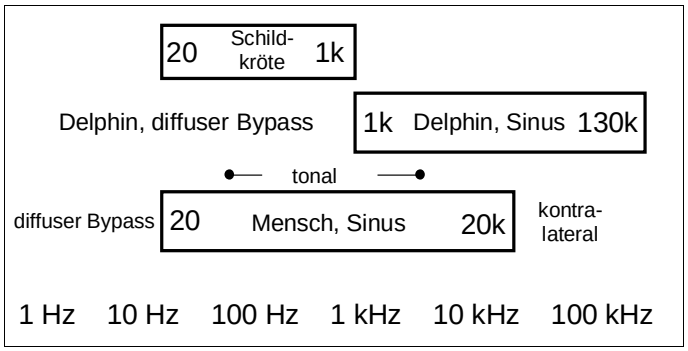

Bild 1: Bereiche diffusen, tonalen und kontralateralen Hörens

Die an der Frequenzgrenze orientierte Abtastfrequenz 44,1 kHz wäre demnach zu niedrig gewählt, wenn nicht die für die Erkennung von Sprache übliche Fensterbreite der Fourieranalyse von typisch 20 ms und auch die visuell auflösbare Pixelgrösse die zeitliche Auflösung des Spektrogramms noch viel stärker beschränken würden. Leonhard.dk [2] weist darauf hin, dass sich u. a. die Laute e, b, und d nur in den ersten Millisekunden voneinander unterscheiden.

2 Laterale Interaktion schon in der Basilarmembran

Von vielen Neuronen weiss man, dass sie im aktiven Zustand eine Erregung konkurrierender Neuronen an ihrer Seite (lateral) hemmen. Für die Haarzellen in der Basilarmembran (BM) hatte man eine ähnliche laterale Hemmung schon zur Erklärung der Resonanzschärfe in Betracht gezogen. Die Haarzellen können sich elektrotonisch [3] durch lokale Anhebung der DC-Komponente sowie fluidmechanisch gegenseitig beeinflussen. Speziell im basalen Abschnitt öffenen sich die Ionenkanäle der als Vorverstärker fungierenden, v-förmig arrangierten äusseren Haarzellen so häufig, dass sich die Wirkungen akkumulieren und für den Steilabfall der Tuningkurve verantwortlich sein können. Plausibel wäre auch, dass die Front vor der von einem einzelnen Schallimpuls bewirkten Wanderwelle aufgesteilt wird und dass die lokale Sprungantwort mit zunächst geringerer Frequenz anschwingt. Möglicherweise erklären sich aus der vermuteten gegenseitigen Beeinflussung der Haarzellen sogar die kritische Bandbreite und ein Teil der schnellen Anpassung an den Pegel des Stimulus über den technisch unerreichten Dynamikbereich 120 dB. Obwohl es kein alternatives rein mechanisches Modell gibt, das alle Phänomene erklären kann, favorisiert man noch immer die schon von Dancer [4] in Frage gestellte Idee wellenförmiger Energieübertragung entlang der Basilarmembran und sieht otoakustische Emissionen sowie Differenztöne als Indizien mechanischer Nichtlinearität an. Fastl hat gezeigt [5], dass Fletcher ausgehend von der Annahme gleicher "akustischer Energie" unzutreffende Werte der kritischen Bandbreite ermittelt hat. Tatsächlich massgebend ist das Zeitmass.

3 Latenz in der dritten Dimension?

Mechanisch-energetische Vorstellungen gemäss dem Ohm'schen Gesetz haben immer wieder dazu verführt unangemessene Mittel der linearen Mathematik wie Fourieranalyse und Autokorrelation auf Systeme aus Nervenzellen anzuwenden. Nicht nur Fletcher [6] suchte in Verzerrungen die Erklärung der von Seebeck entdeckten Hörbarkeit real fehlender Grundwellen. Seit Beweise gegen die Autokorrelations-theorien vorliegen [7] und neuronale Netze mit zeitlich kodierten Signalen und dynamischen Synapsen das natürliche Gehör erstmals in der sprecherunabhängigen Spracherkennung übertroffen haben [8], besteht erneut Veranlassung herauszufinden, was wirklich passiert.

Im Nucleus cochlearis (CN) gabelt sich der Hörnerv in drei Äste. Die in der BM eindimensional angelegte tonotopische Ordnung nach der Best-Frequenz (BF) bleibt jedoch offensichtlich bis in den Cortex mehr oder weniger als eine der Dimensionen erhalten. Langner ordnet einer der beiden anderen Dimensionen des Raums im Colliculus inferior (IC) eine periodotopische Ordnung nach der besten Modulationsfrequenz (BMF) zu und macht diese für die Tonhöhe verantwortlich, im Unterschied zum Klang, den er dem Frequenz-Spektrum zuschreibt, das entlang der BM tonotopisch manifestiert ist [9]. Auch dieser bedeutende Fortschritt gegenüber der klassischen, rein spektralen Betrachtung löst die Frage nach der Funktion des Gehörs noch nicht erschöpfend. Die beteiligten Neuronen und ihre Verschaltung sind vielfältig und manch psychoakustischer Effekt, den man über Templates zu erklären versucht, mag eine simplere Ursache haben. Beispielsweise entspricht die Oktav-Vergrösserung [10] einer Verkürzung des Intervalls um etwa 15 µs, als ob dieses zwischen fallender und steigender Flanke zweier so schmaler Aktionspotentiale gemessen worden wäre. Möglicherweise ist in der dritten Dimension von Teilen des IC die mit CF einhergehende cochleare Latenz kodiert. Die Gehirnfunktion ist u. a. deshalb schwer zu verstehen weil z. B. zum integralen Parameter Tonhöhe viele Neuronen zusammen beitragen, die nicht unbedingt räumlich benachbart sein müssen. Neuronen triggern dann, wenn an ihren Wurzeln (Dendriten) genügend erregende Potentiale zusammentreffen (koinzidieren).

4 Kontralaterale Koinzidenz

Vom CN bis zum Cortex wechseln etliche auf- und auch einige absteigende Projektionen von der linken zur rechten Hirnhälfte hinüber bzw. umgekehrt. Wenngleich bisher noch keine Idee zur Abbildung binauraler Aspekte im Spektrogramm bekannt ist, wird nachfolgend beispielhaft jene Verbindung zwischen beiden Ohren betrachtet, der man die azimutale Ortung von Quellen niedriger Frequenz oder kurzzeitigen Schalls zuschreibt. Sie kodiert in beiden medialen oberen Oliven (MSO) den Versatz zwischen den Mustern auf der BM des linken und des rechten Ohrs um maximal etwa 650 µs mittels zeitlicher Koinzidenz in eine Ortskoordinate. Als notwendige Verzögerungsleitung steht dabei jeweils die BM zur Verfügung. Das kontralaterale und das ipsilaterale Signal treffen sich in von beiden Ohren erregten Zellen, und zwar bei geringem Versatz im tonotopisch hier auffällig breiten Bereich [1] niedriger BF. Zur exzellenten Zeitauflösung bis herunter zu 4 µs trägt also neben den Eigenschaften der Haarzellen, des Hörnervs und der sphärischen Büschelzellen auch eine Spreizung der Tonotopie im von linearen Spektrogrammen sogar komprimierten Niederfrequenzbereich bei.

5 Tonale Koinzidenz

Man kann sich vorstellen, ein neuraler Baustein ist auf ein bestimmtes Intervall der Stimulation bzw. Vielfache davon abgestimmt. Dies bedeutet auch, dass an seinem Ausgang Folgen von Impulsen (Aktionspotentialen) erscheinen, die mit dem Zeitraster seines Eingangssignals synchronisiert sind. Nur deshalb kann beispielsweise eine Oktave als Gleichklang wahrgenommen werden. Man nennt es zwar spektrale Kodierung, wenn beispielsweise die Töne a4 (440 Hz) und e4 (330 Hz) tonotopisch lokalisiert sind. Das für die Harmonie entscheidende Merkmal ist jedoch gar keine dieser Frequenzen, auch nicht ihr Kehrwert, sondern die sich mit der wahrgenommenen Grundperiode a2 (10ms/1,1 = 4/440Hz = 3/330Hz) wiederholende zeitliche Koinzidenz. Man hört diese physikalisch nicht existente "Grundwelle" freilich nur unter der Voraussetzung, dass a4 und e4 kohärent sind, also bei natürlichen Geräuschen den gleichen Ursprung haben. Nur kohärente Bestandteile erkennt das Gehör zusammengehörig als Vokal. Andere werden wie falsche Formanten wahrgenommen. Noch ist nicht ganz geklärt, an welcher Stelle des Gehirns diese Koinzidenz detektiert wird. Nuclei unterhalb des IC dürften ausscheiden. Als skalierbaren Ton hört man den kleinsten gemeinsamen Nenner auch nur dann, wenn er innerhalb des Hörbereichs liegt. Dies ist sowohl psychoakustisch als auch neurophysiologisch plausibel. Weil also die sogenannte spektrale Ordnung tatsächlich auf zeitlicher Koinzidenz beruht, darf man nicht erwarten, dass die im Spektrogramm veranschaulichte Analyse dem Gehör gerecht wird. Das Heraushören spektraler Komponenten ist bei Harmonie recht zweifelhaft. Man sieht im Spektrum verwirrend viele parallele Linien, die sich dem einheitlichen Höreindruck kaum zuordnen lassen. Korrelate physikalisch fehlender Grundfrequenzen, die man tatsächlich hört, vermisst man nicht nur im Spektrogramm. Von virtueller Tönhöhe ausserhalb übertragener Bandbreite profitieren beispielsweise am Telefon gut verständliche tiefe Männerstimmen.

Einerseits kann es sinnvoll sein, Spektrogramme um die Tonhöhe bzw. Grundfrequenz zu ergänzen. Dies wurde erfolgreich versucht. Andererseits ist für die kognitive Verarbeitung akustischer Muster das bisher fast ausschliesslich betrachtete Spektrum absoluter Frequenzen sicherlich ebenso nebensächlich wie die Stimmlage eines Sprechers für seine Aussage. Unabhängig von neurophysiologischer Realität wurde deshalb nach Möglichkeiten gesucht, in einer Zusatzdimension des Spektrogramms auch die Harmonie aufzuzeigen. Die Zyklizität der zwölf Halbtöne c bis h im Tonklassenkreis wird hierfür auf die sechs Farben rot bis violett des Farbkreises abgebildet.

6 Diffuser Bypass (einschliesslich modulation transfer)

Die sogenannte Phasenkopplung gilt als Grundlage einer zusätzlichen, zeitlich genannten Wahrnehmung, welche quasi vorbei an der cochlearen Spektralanalyse vor allem von der grossen Zahl jener Fasern des Hörnervs vermittelt wird, die höheren CFs zugehören und die in rascher zeitlicher Folge aber strenggenommen überhaupt nicht gleichzeitig feuern. Deren Überlappung scheint man immer nur dann vordergründig als Koinzidenz wahrzunehmen, wenn die stärkere tonale Koinzidenz nicht ausreichend funktioniert. Dies betrifft vor allem Periodizitäten nahe der unteren Frequenz-Grenze hörbarer Sinustöne oder darunter, schliesst aber auch die generelle absolute Skalierbarkeit der Tonhöhe ein, in Ergänzung zur mehrdeutigen tonalen Koinzidenz. Man hat empirisch den fragwürdigen Eindruck gewonnen, dass die Wahrnehmung der Phasenkopplung der zeitlichen Hüllkurve des originalen Stimulus zuzuschreiben ist. Wenn behauptet wird, die zeitliche Struktur bleibt erhalten, so ist dies auch nicht ganz korrekt. Tatsächlich wird sie infolge Verzögerung entlang der BM deformiert. Beispielsweise evoziert ein einzelner Schallimpuls als Sprungantwort ein in Form von Wanderwellen über die BM laufendes Muster von dem sich nur die ersten 2 bis 3 Amplituden als Folgen von Aktionspotentialen mit stetig zunehmender Verzögerung abbilden. Dem impulsförmigen Stimulus wäre keine zeitliche Hüllkurve angemessen. Stellvertretend sei die Summe über alle Nervenimpulse betrachtet, die gleichzeitig in ein an der Breite der Aktionspotentiale orientiertes schmales Zeitfenster fallen. Bild 2 lässt erkennen, wie die Aktivität abklingt, während die Wanderwelle zu niedriger CF läuft. Dieser Verlauf entspricht einem Integral und ist in anderen Fällen der zeitlichen Hüllkurve ziemlich ähnlich.

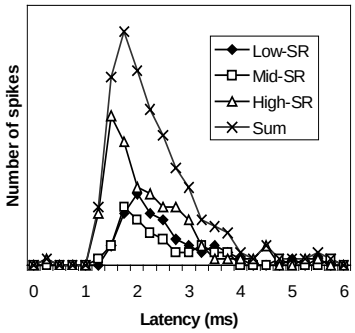

Bild 2: Anzahlen von Zellen die nach einem kurzen Klick mit niedriger, mittlerer und hoher Spontanrate (SR) sowie insgesamt jeweils fast gleichzeitig aktiv waren. Messung in 334 Fasern der Hörnerven von acht Katzen. Daten von Cai [11] sind in Abhängigkeit von der Laufzeit (Latenz) auf der Basilarmembran dargestellt.

Abgesehen von ihrer möglicherweise zu geringer Zeitauflösung sind Spektrogramme grundsätzlich geeignet, derartige Pulsationen bzw. Modulationen in zeitlicher Dimension aufzuzeigen. Im Gehirn gibt es offenbar auf Modulationsfrequenzen abgestimmte Neuronen. Ihre Funktion wird vom Modulations-Spektrogramm [12] nachgebildet.

7 Sensitivität gegenüber schnellen Änderungen

Prinzipbedingt ist die tonale Koinzidenz im stationären Fall genauso phasentaub wie die mechanische Spektralanalyse auf der BM. Die folgende Vermutung [13] trifft für an- und abklingende Hüllkurven bei hinreichend niedriger Folgefrequenz zunächst zu: "Wenn Veränderungen im Phasenspektrum zu einer Veränderung der Hüllkurve des Signals führen, dann ist die Differenz hörbar". Beim abklingenden Abwärts-Chirp ist der Anstieg steiler, die Koinzidenz stärker ausgeprägt und folglich hört man einen schärferen Klicklaut. Das Gehör als Kantenfilter bewertet Onsets stärker als Offsets. Dass diese Eigenschaft nicht an die Form der "Hüllkurve" bzw. den diffusen Bypass gebunden ist, wurde mit einem Sinussignal bewiesen, das im Nulldurchgang einen winzigen, kaum sichtbaren Versatz erhielt. Bei einmaligem Versatz hört man einen Klick, bei periodisch wiederholtem ändert sich der Ton polaritätsabhängig.

Während man also den stationären Einfluss von Phase und Polarität nur bei niedriger Pulsationsfrequenz, typisch um 100 Hz, überhaupt bemerkt, vermelden die Neuronen schnelle Abweichungen extrem empfindlich. Der Vorzustand mag dabei auf verschiedene Weise gespeichert sein, zunächst mechanisch in der Trägheit der BM, dort zugleich auch elektrochemisch, dann in der Abfolge hemmender Wirkung zuerst stimulierter Neuronen höherer CF auf parallele, später angesprochenen, sowie insbesondere im Verhalten von Onset-Choppern. Umfang und Kompliziertheit von Nerven-Projektionen, die als hemmend identifiziert sind, belegen, dass man diesen Aspekt des Hörvorgangs bisher unterschätzt. Die Spektralanalyse wäre kein guter Ratgeber. Massgebend sind Änderungen der Abstände zwischen den Aktionspotentialen, unabhängig davon, ob sie ursächlich Änderungen der Phase oder der Frequenz entsprechen. Vermutlich treten in vielen Fällen jene Veränderungen besonders hervor, welche unmittelbar die tonale Koinzidenz betreffen, Farbkontraste gewissermassen. Sie können durch Blinken hervorgehoben werden. Ein besseres Verständnis dynamischer Signalverarbeitung bis hin zur Hemisphären-Zuordnung wird zur Berücksichtigung der schnellen Verschiebung von Formanten im Spektrogramm bzw. Preprozessor und letzlich zur besseren Erkennung von Schallmustern beitragen.

Literatur

1 P. Buser, M. Imbert: 1992 Audition. MIT Press.
2 F. U. Leonhard: 1998 IEEE ICASP, Seattle.
3 A. J. Hudspeth: 2000 Persönliche Mitteilung.
4 A. Dancer et al.: 1997 Diversity in Aud. Mech. Singapore, 340-346.
5 E. Zwicker, H. Fastl: 1999 Psychoacoustics. 2nd ed. Springer.
6 R. M. Warren: 1999 Auditory Perception. Cambridge Univ. Press.
7 Ch. Kaernbach, L. Demany: 1998 JASA 104 (4), 2298-2306.
8 M. F. Yeckel, T. W. Berger: J. 1998 Neuroscience 18, 438-450.
9 G. Langner: 1997 J. New Music Res. 26, 116-132.
10 M. McKinney, B. Delgutte: 1999 JASA 106 (5), 2679-2692.
11 Y. Kai: 1995 Ph.D. thesis, Univ. Wisconsin-Madison.
12 S. Greenberg, B. Kingsbury: 1997 IEEE ICASP Munich, 1647-50.
13 S. Tempelhaar: 1996 Signal Proc., Speech & Music. Swets & Zeitl.