

## Aufmerksam hören

Jens Blauert<sup>1</sup>, Thomas Walther<sup>2</sup>

<sup>1,2</sup> *Institut für Kommunikationsakustik, Ruhr-Universität Bochum, 44780 Bochum, jens.blauert@rub.de*

### Einleitung

*Aufmerksamkeit* beim Hören (engl.: *auditory attention*) ist ein Phänomen, mit dem sich traditionell insbesondere die sensorische und die kognitive Psychologie befasst haben – vergl. z.B. Nobre & Kastner (2014), Carrasco (2011), Jones & Yee, (1993), sowie die Übersichten von Spence & Santangelo (2010) und Hafter *et al.* (2008). Im Zuge des Fortschrittes der modernen Technologien gewinnt aber Aufmerksamkeit zunehmend an praktischer Bedeutung, z.B. in der Robotik, der Hörgerätetechnik, für intelligente Erkennungs-, Identifikations- und Beurteilungssysteme aller Art, wie z.B. Spracherkenner, Dialogsysteme, Sound-Quality-Schätzer. Im Folgenden werden einige Grundlagen von Aufmerksamkeit beim Hören erörtert, wobei deren potentielle technische Anwendbarkeit im Vordergrund steht. Dadurch wird ein spezieller Aspekt hervorgehoben, nämlich der folgende:

Aufmerksamkeit ist u.a. ein Vorgang, bei dem die Komplexität der Szenenanalyse sich dadurch reduziert, dass eine situationsbezogene Fokussierung auf die wesentlichen Objekte oder deren Merkmale stattfindet (Engelke *et al.* 2011). Hierdurch wird in der Regel die Leistungsfähigkeit eines System – sei es biologisch oder technisch – in Bezug auf vorgegebene Aufgaben erhöht. Man erkennt schnell, dass so ein Verhalten Rückkopplung von „höheren“ zu „niedrigen“ Verarbeitungsstufen innerhalb des Systems erfordert. Je nach den Ausgangspunkten der beteiligten Rückkopplungspfade kann man Aufmerksamkeitsvorgänge in *reflexive* und *reflektive* einordnen. Es sei aber angemerkt, dass es sich dabei um eine operationale Unterscheidung handelt und dass diese beiden Kategorien nicht notwendigerweise unabhängig voneinander sind (Spence 2007).

### Reflexive Aufmerksamkeit

Diese Art von Aufmerksamkeit folgt auf „einfache“ Merkmale des Wahrgenommenen (*primitive perceptual cues*), ohne dass eine kognitive Verarbeitung stattfindet. Die Aufmerksamkeitsreaktion besteht in einer schnellen Hinwendung zu Objekten oder deren Positionen im Wahrnehmungsraum innerhalb vergleichsweise kurzer Reaktionszeit. Dies geschieht unabhängig von einer vorgegebenen speziellen Aufgabe. Auch bedarf es zur reflexiven Aufmerksamkeitsreaktion keines speziellen Trainings, und die Reaktion kann nicht umtrainiert werden.

Es wird vermutet, dass die zu reflexiver Aufmerksamkeit führenden Einsatzsignale (*cues*) nicht ignoriert werden können. Sie ereignen sich vorwiegend nach überraschenden Änderungen im perzeptiven Umfeld, z.B. nach Positionsänderungen von Objekten, plötzlichem Auftauchen unbekannter Geräusche, Änderungen von Tonhöhe, Klangfarbe oder Lautheit. Die reflexiven Reaktionen klingen innerhalb kurzer Zeit wieder ab (weniger als 1s). Dies erinnert an Schreckreaktionen. Dazu passt, dass nach nachfolgenden gleichartigen Einsatzsignalen die Reaktion verzögert oder gar nicht

eintritt. Falls die reflektive Reaktion nicht zu einer besseren Wahrnehmung von deren „Ursache“ beiträgt, kann sogar eine zeitlich begrenzte Unorientiertheit eintreten – was de facto das Gegenteil vom Zustand des Aufmerksamkeit ist.

### Reflektive Aufmerksamkeit

Reflektive Aufmerksamkeit ist geplant und zielorientiert. Die Reaktionen können durch Lernen und Training modifiziert werden. Sie erfordern folgerichtig Rückkopplung von kognitiven Verarbeitungsstufen. Je nach der Art der reflektiven Reaktionen müssen zu ihrem Verständnis vielfältige Einsatzsignale sowie Zustandsparameter des Systems in Betracht gezogen werden. Diese können unterschiedlichste Abstraktionsgrade aufweisen, beginnend mit einfachen (*primitive*) psychoakustischen Merkmalen wie Lautheit und Klangfarbe, über komplexere perzeptiven Konstruktionen wie Lästigkeit und Sound-Qualität bis hin zu nicht-auditiven Randbedingungen wie kreuzmodalen Einflüssen (z.B. vom Gesichtssinn), dem augenblicklichen emotionalen und aktionalen Zustand des Systems (inklusive laufendender Kommunikationsvorgänge mit anderen Systemen), dem Wissen über das Umfeld und dessen Geschichte, speziellem Expertenwissen, und vielem Anderen mehr – vergl. Blauert & Jekosch (2012) Es sind diese reflektiven Aufmerksamkeitsvorgänge, die wegen ihres technischen Anwendungspotentials z. Zt. besonderes Interesse hervorrufen.

### Empirische Grundlagen – eine Auswahl

#### Aufmerksamkeit bei Detektion von Schmalbandsignalen

Die Detektion von Schallsignalen betrifft die Fähigkeit, gesuchte Signalanteile aus dem störenden Hintergrund heraus zu heben. Die meisten Erklärungsversuche dazu berücksichtigen, dass das Gehör eine Zerlegung in Frequenzbänder (sog. Frequenzgruppen, engl.: *critical bands*) vornimmt (Fletcher 1940; Zwicker 1961). Zuhörer entdecken schmalbandige Signale leichter, wenn sie in eines dieser Bänder fallen. Die Entdeckungsleistung steigt, wenn die Zuhörer darüber informiert sind, in welchem der Bänder das Signal zu erwarten ist; ihre Aufmerksamkeit also auf dieses Band gelenkt ist (Greene 1960, Scharf 1970). Erscheint das Signal dann aber nicht in dem erwarteten Frequenzband, sinkt seine Entdeckbarkeit drastisch ab (Green & Swets 1966).

Bei Klängen, die spektral breiter sind als eine Frequenzgruppe, berücksichtigen die Zuhörer bei der Entdeckung zwar auch solche Anteile, die nicht in denjenigen Bändern positioniert sind, auf die die Aufmerksamkeit gerade gerichtet ist. Laut Dai *et al.* (1991) werden diese Anteile vom Gehör bei der weiteren Verarbeitung allerdings kaum mehr beachtet. Man kann die Erkennungsgewinne aus zusätzlichen Bändern jedoch auch als eine vollzogene Verbreiterung des dominanten Spektralbereiches deuten. Dann ergibt sich die Frage, auf welcher Grundlage das Gehör entscheidet, den beobachteten Spektralbereich zu verbreitern. Hafter *et al.* (2008) vermuten, dass dies geschieht, wenn ein abstrakteres

als die einfachen (primitiven) Merkmale erkannt wird, z.B. eine Residuum-Tönhöhe oder eine klare Hörereignisposition. Auch solche komplexeren Merkmale sind also durchaus in der Lage, Aufmerksamkeit auf sich zu ziehen.

### Aufmerksamkeit beim Hören von Musik

Aufmerksamkeit spielt nicht nur bei der Entdeckung von Signalen in Störgeräuschen eine Rolle, sondern kann sich auf beliebige Hörereignismerkmale richten, seien sie nun gestört oder klar hörbar. Weiterhin können damit assoziierte konzeptionelle Merkmale Aufmerksamkeitsziele sein. Bei Musik kann man sich z.B. außer auf Rhythmus, Tonhöhe, Lautheit, Rauigkeit, Klangfarbe, usw., auch auf Merkmale wie Melodie, Klangqualität, Genre, Aufführungsqualität oder gar auf den emotionalen Eindruck (*impact*) fokussieren.

Wenn Zuhörer sich auf einfache (primitive) Hörereignismerkmale konzentrieren, begeben sie sich in einen analytischen Hörmodus, in welchem sie versuchen, die Hörereignismerkmale selektiv wahrzunehmen. Auditive Objekte werden dann nicht in ihrer Ganzheit beachtet. Wenn es aber z.B. die Aufgabe ist, die Sound-Qualität einer musikalischen Szene zu beurteilen, geschieht dies in einem integrativen Hörmodus. Ersterer ist sicherlich eher reflexiv und letzterer reflektiv. Jones & Yee (1993) unterscheiden in diesem Zusammenhang zwischen *integrative* und *selective attention*.

Die musikalischen Signale dienen u.a. als Träger von Bedeutung. Wenn die Kommunikation von Bedeutung im Vordergrund des Interesses steht, kommt es nicht in erster Linie darauf an, wie ein Musiksignal klingt, sondern darauf, in wie weit es in der Lage ist, Bedeutung zu transportieren. Die Kommunikationsqualität von Musiksignalen ist also nicht gleichbedeutend mit deren Klangqualität – vergl. Jekosch 2005, Blauert & Jekosch 2012, Raake & Blauert 2013.

Übrigens sind Zuhörer, die sich voll auf spezielle Hörmerkmale konzentrieren, oftmals taub für alle anderen. Akustische Berater, die gewohnt sind, genau auf die Raumeigenschaften zu hören, achten z.B. oft nicht auf die Musik, die gerade spielt wird. Dieser aufmerksamkeitsbedingte Ablenkungseffekt wird auch *attentional amnesia* genannt (Wolfe 1999, Gregg & Samuel 2008, Shinn-Cunningham 2008).

### Aufmerksamkeit beim Hören von Sprache

Gesprochene Sprache ist das prominente Medium zwischenmenschlicher Kommunikation. Somit sind Aufmerksamkeitseffekte von ganz besonderer Bedeutung für die Sprachsignalverarbeitung. Selektive Aufmerksamkeit auf Sprachsignale ist besonders dann von Bedeutung für die Sprachverstehbarkeit, wenn z.B. Lärm, Nachhall, lineare oder nicht-lineare Verzerrungen, konkurrierende Sprecher, oder auch störende Einflüsse aus anderen Sinnesgebieten präsent sind. Vergleiche z.B. Moore et al. (2010), Darwin (2007/2010), Assmann & Summerfield (2004) zu diesem Problemkreis.

Eine neue Sichtweise auf diese Thematik ergab sich durch die Pionierarbeiten von Cherry (1953, 1954), der auch den Ausdruck „Cocktail-Party-Effekt“ prägte, und zwar für das Phänomen, dass es möglich ist, sich in einer Gruppe durcheinander redender Sprecher auf einen davon zu konzentrieren und diesen aus dem allgemeinen Sprachgemisch heraus

zu hören. Von vielen zu diesem Thema durchgeführten Untersuchungen ist die folgende besonders aufschlussreich.

Man spielt Zuhörern mit Kopfhörern Sprache von zwei unterschiedlichen Sprechern vor, und zwar diejenige des einen auf das eine Ohr und diejenige des anderen auf das andere Ohr. Die Zuhörer müssen dann laut nachsprechen was der eine der beiden Sprecher sagte. Werden sie dann nach dem Experiment gefragt, was denn der jeweils andere Sprecher gesagt hat, sind sie dazu nicht in der Lage. Sie haben allenfalls bemerkt, ob der andere Sprecher männlich oder weiblich ist, oder dass er während des Experimentes ausgetauscht oder durch einen Sinuston ersetzt worden ist. In anderen Worten, sie können sich hinsichtlich der Bedeutung des Gesagten nur auf einen Sprecher konzentrieren. Von dem jeweils anderen nehmen sie bestenfalls subsemantische Sprachmerkmale wahr. Aber es gibt Ausnahmen: Wenn z.B. der andere Sprecher mehrfach den Namen eines Zuhörers ausspricht, wird das von diesem Zuhörer durchaus bemerkt. Offensichtlich werden also sowohl einfache als auch komplexe Sprachsignalmerkmale von den Zuhörern verarbeitet. In beiden Fällen ist Aufmerksamkeit mit im Spiel. Ob allerdings die Verarbeitung der unterschiedlichen Merkmalklassen parallel oder sequenziell erfolgt, lassen die hier geschilderten Versuchsergebnisse allein nicht erkennen,

Bezüglich Aufmerksamkeit bei Sprechsprache bleibt somit eine wichtige Frage offen: Wie erfolgt die Fokussierung auf bestimmte Sprecher bzw. deren Sprachsignale im Einzelnen? Eine Hypothese dazu lautet wie folgt. Das auditorische System wirft kurze, flüchtige Blicke (*glimpses*) auf die ankommenden Sprachsignale und wählt für die weitere Verarbeitung solche Segmente aus, die eine in Hinblick auf spezifische Sprechermerkmale möglichst geringe statistische Unsicherheit aufweisen (Miller & Licklider 1950, Bregman 1991, Cooke 2003). Diese Segmente dienen als Grundlage für die weitere Verarbeitung. So elegant diese Hypothese auch sein mag, sie erklärt nicht, wie auf einer noch recht niedrigen Verarbeitungsstufe bereits erkannt werden kann, welche *glimpses* für die höheren (kognitiven) Verarbeitungsstufen signifikante Information liefern, z.B. bei der Objektbildung oder bei der Zuordnung von Bedeutung. *Glimpsing* ist deshalb eher als ein Mechanismus reflexiver Aufmerksamkeit einzustufen, bei dem u.a. *Gestalt*-Erkennung eine Rolle spielen mag – siehe nächster Abschnitt.

### Objektbildung und Gestaltregeln

In Zuge der Bildung von Hörobjekten und daraus zusammengesetzten Hör Szenen sind Routinen erforderlich, die diejenigen einfachen (*primitive*) auditiven Merkmale identifizieren und verbinden, die für die Bildung von Hörobjekten erforderlich sind. Dies sind in der Regel Merkmale, die den Schallsignalen einer jeweils individuellen Quelle zuzuordnen sind. Dieser Prozess wird Gruppierung (*grouping*) genannt (Bregman 1991, 1993, Warren 1982). Man unterscheidet zwei Arten von Gruppierungsroutinen, nämlich:

- (1) *Einfache Gruppierungsroutinen*: Diese sind sozusagen fest verdrahtet und arbeiten schnell und unbewusst. Man kann sie nicht modifizieren oder manipulieren.

- (2) *Erlernte Gruppierungsroutinen*: Diese werden in der kognitiven Psychologie *Schemata* genannt. Schemata repräsentieren kognitive Prozesse und optimieren sich selbsttätig im Zuge von allfälligen Erfahrungen mit der Umwelt. Sie lassen sich umtrainieren und lernen ggf. auch aus Anweisungen und Erläuterungen.

Die einfachen Gruppierungsroutinen arbeiten nach Regeln, die die Gestalt-Psychologie entdeckt hat, u.a. den Regeln von Prägnanz, Nähe, Ähnlichkeit, Figur und Hintergrund, gemeinsamem Schicksal, Erfahrung, Geschlossenheit, Einfachheit, Kontinuität – vergl. dazu z.B. Von Ehrenfels, (Neuaufgabe 1990) und die Übersicht von Jekosch (2005).

Die Schema-basierten Routinen repräsentieren kognitive Prozesse und sind somit komplexer als die einfachen. Dennoch arbeiten sie schnell und zielorientiert. Sie sind typischerweise aufgabenspezifisch und auf ihre spezielle Aufgabe optimiert (vergleichbar mit Apps in Smartphones). Sie stellen somit effiziente Routinen dar, die in ihrem speziellen Umfeld plausible Resultate liefern.

Offensichtlich ähneln die beiden Arten von Gruppierungsroutinen den beiden Arten von Aufmerksamkeit (*reflexiv* und *reflektiv*) die zu Anfang dieses Artikels unterschieden wurden. Die unterliegenden Prinzipien sind jeweils gleich, und es gilt auch, dass jeweils beide Arten konfluent sind.

### Technische Realisierung von Aufmerksamkeit

Ansätze zur technischen Realisierung von auditiven Aufmerksamkeitsvorgängen sind noch selten. Sie basieren fast ausschließlich auf Software-Algorithmen. Im Folgenden wird eine Auswahl von Ansätzen beschrieben, die auf das kürzlich beendete EU-Projekt „TWO!EARS“ zurückgeht. (TWO!EARS, 2017). Alle Resultate inkl. der Software dieses Projektes sind „open source“ und somit frei verfügbar.

#### Den Kopf drehen

Aufmerksamkeit kann sich durch Hinwenden des Kopfes zu Schallquelle ausdrücken. Damit wird die Quelle in das Blickfeld gerückt (so-called *turn-to reflex*). Diese Aktion wird aber fälschlich als reflexiv aufgefasst, denn sie kann willentlich modifiziert werden. Blinde drehen den Kopf z.B. oft bewusst etwas zur Seite in die Richtung des besten Hörens. Auch andere Kopfdrehbewegungen können je nach Situation sinnvoll sein. Alle technischen Mittel, um solche zu simulieren, stehen nun zur Verfügung, u.a. ein Kunstkopf mit steuerbarer Kopfdrehvorrichtung, weiterhin Algorithmen zur zuverlässigen Bestimmung der Richtungen, in denen sich Schallquellen befinden.

#### Abschätzen welche Schallquellen wichtig sind

In einem akustischen Umfeld, in dem ständig Schallquellen erscheinen, sich bewegen und wieder verschwinden, wäre es uneffektiv, alle diese Quellen ständig durch Hinwenden zu verfolgen. Es wurde deshalb ein Algorithmus entwickelt, der abschätzt, ob eine neu auftauchende Quelle wichtige neue Information bereit stellt. Dies ist z.B. in der Regel nicht der Fall, wenn sie zum wiederholten Male auftaucht oder aus anderen Gründen als bekannt einzustufen ist. Als ein Maß für die Wichtigkeit einer Schallquelle kann gelten, ob sie in Bezug auf das Umfeld kongruent oder inkongruent erscheint.

Im letzteren Fall wird sie als überwachungswürdig eingestuft und erfährt konsequenterweise Aufmerksamkeit. Zum Beispiel wird dann eine Hinwende-Aktion eingeleitet.

#### Die Ohren „spitzen“

Mütter sind bekanntlich gut darin, das Schreien ihres Kindes aus einem Stimm- und Lärmgewirr heraus zu hören (ein spezieller Fall des *cocktail-party effects*). Dies erfordert aber hohe Aufmerksamkeit. Man sagt auch, man muss dazu die Ohren „spitzen“. Das TWO!EARS-System wurde darauf trainiert, das menschliche Hörvermögen auch in diesem Bereich zu simulieren. Es ist nun in der Lage, ein Schallsignal, welches es kennt, besser aus einem Signalgewirr heraus zu hören und außerdem die Quellposition besser zu bestimmen. Hierbei werden u.a. geeignete Kopfdrehungen vorgenommen. Bei zwei konkurrierenden Schallquellen, die beide Aufmerksamkeit erfordern, nimmt der Kopf ggf. die Position ein, in der sich die beiden Schallsignale auditiv bestmöglich trennen lassen. Auch hierbei ist es von Vorteil, wenn das System auf die beteiligten Signale trainiert ist – zumindest aber auf deren Signalklassen.

#### Schallquellen in Räumen orten

In einem Raum mit schallreflektierenden Wänden treffen bei den Zuhörern nicht nur die von den Quellen stammenden Direktschalle ein sondern zusätzlich, allerdings zeitlich verzögert, die von den Wänden reflektierten Schalle, die sog. Rückwürfe. Obwohl die Rückwürfe aus anderen Richtungen als der Direktschall kommen, besteht bei den Zuhörern in der Regel kein Zweifel bezüglich der Richtungen in den die Schallquellen stehen. Dieses Phänomen heißt *Präzedenzeffekt*. Bei der Bildung der Höreignisrichtungen vernachlässigt das Gehör also die Schalleinfallrichtungen der Rückwürfe, obwohl es die Existenz von Rückwürfen durchaus wahrnimmt – u.a. durch Halligkeit, Klangfarbenänderung und erhöhte Lautheit. In TWO!EARS wurde nunmehr ein neuartiger Algorithmus entwickelt, der den Präzedenzeffekt überzeugend simuliert. Neu gegenüber früheren Ansätzen ist dabei, dass die Steuerinformation für die Simulation aus den laufenden Signalen gewonnen wird, z.B. aus der Musik die während eines laufenden Konzertes erklingt. Das Verfahren ist so erweiterbar, das es nicht nur die Quellenrichtungen ermittelt, sondern auch die aus diesen Richtungen fortlaufend eintreffenden Schalle getrennt zur Verfügung stellt (*auditory-stream segregation*). Weiterhin ist eine Verminderung des in diesen Signalen enthaltenen Nachhalls möglich (*de-reverberation*).

#### Rettungsaktionen nach Gehör leiten

Ob ein künstliches System in der Lage ist, eine akustische Szene zu analysieren und zu interpretieren, kann man z.B. damit nachweisen, dass man ihm eine Aufgabe zuteilt, die es nur sinnvoll ausführen kann, wenn es Szene und Aufgabe „verstanden“ hat. In TWO!EARS soll das System, welches sich auf einer Roboterplattform bewegen kann, nach einem plötzlich ausgebrochenen Brand gefährdete Personen in Sicherheit bringen. Die Reihenfolge der Rettung der einzelnen Personen soll dem Gefährdungsgrad der einzelnen Subjekte angepasst sein. Zu Orientierung dienen (emulierte) auditive Objekte, die durch perzeptive Merkmale definiert sind. Bei Bedarf kann das System zusätzlich eine simulierte

Kamera aktivieren, um liegende von stehenden Personen zu unterscheiden. Das TWO!EARS-System bewältigt die gestellte Aufgabe, indem der Roboter zunächst die Positionen aller aktiven Schallquellen (simulierte Personen) kartographiert. Dabei werden auch die spezifischen auditiven Merkmale der einzelnen Quellen erfasst. Sobald die Situation in ein S&R-Szenario übergeht (Ausbruch eines Feuers im zu überwachenden Gebiet) löst der Roboter Alarm aus und führt unter Berücksichtigung der aktuellen individuellen Gefährdungsgrade der erfassten Schallquellen sukzessiv die Rettung der durch diese Schallquellen repräsentierten Personen durch.

## Danksagung

Die Vorarbeiten zu diesem Artikel wurden durch das EU-Projekt TWO!EARS (FP7-ICT-2013-C) unterstützt. Innerhalb dieses Projektes wurde der Themenbereich „Aufmerksamkeit“ im Arbeitspaket „Auditory Feedback“ bearbeitet – und zwar von folgenden Personen: Argentieri, S., Blauert, J., Braasch, J., Bustamante, G., Cohen-L'hyver, B., Danès, P., Forgue, Th., Guo, Y., Kashef, Y., Ma, N., May, T., Mohr, J., Podlubne, A., Kim, Ch., Walther, Th. und Wierstorf, H.

## Literatur

- Assmann, P. F. & Summerfield, Q. (2004) "The perception of speech under adverse conditions," in: *Speech processing in the auditory system*, edited by S. Greenberg, W. Ainthworth, A. Popper & R.R. Fay, Springer, New York, 231–308
- Blauert, J. & Jekosch, U. (2012) "A layer model of sound quality," *J. Audio Engr. Soc.* 60, 4–12
- Bregman, A. S. (1991) "Using quick glimpses to decompose mixtures," in: *Music, language, speech and brain*, edited by J. Sundberg, L. Nord, and R. Carlson, MacMillan, London, 284–289
- Bregman, A.S. (1993) "Auditory scene analysis: Hearing in complex environments," in: *Thinking in s.Sound: the cognitive psychology of human audition*, edited by S. McAdams and E. Bigand, Clarendon Press, Oxford, chap. 2, 10–36
- Carrasco, M. (2011), "Visual attention: The past 25 years" *Vision Res.* 51, 1484–1525.
- Cherry, E. C. (1953) "Some experiments upon the recognition of speech with one and two ears," *J. Acoust. Soc. Am.* 25, 975–979
- Cherry, E. C. (1954) "Some further experiments upon the recognition of speech with one and two ears," *J. Acoust. Soc. Am.* 26, 554–559
- Dai, P., Scharf, B. & Buus, S. (1991) "Effective attenuation of signals in noise under focused attention," *J. Acoust. Soc. Am.* 89, 2837–2842
- Darwin, C. J. (2007/2010) "Listening to speech in the presence of other sounds," in: *The perception of speech: from sound to meaning*, edited by B. Moore, L. Tyler & W. Marslen-Wilson, Oxford univ. press, Oxford, chap. 7, 151–170, published online first (2007)
- Engelke, U.H.K., Zepernick, H.-J. & Ndjiki-Nya, P. (2011) "Visual attention," *IEEE Sign. Proc. Mag.*, 50–59
- Fletcher, H. (1940) "Auditory pattern," *Rev. Modern Physics* 12, 47–65
- Green, D. & Swets, J. A. (1966) *Signal detection theory and psychophysics*, John Wiley & Sons, New York
- Greene (1960) "Psychoacoustics and detection theory," *J. Acoust. Soc. Am.* 32, 1189–1203
- Gregg, M. K. & Samuel, A. G. (2008) "Attention and the organizational properties of sound," *J. Exp. Psychol.: Human Perception and Performance* 5, 168–175
- Haftner, E. R., Sarampalis, A. & Loui, P. (2008) "Auditory attention and filters," in: *Auditory perception of sound sources*, edited by W. A. Yost, A. N. Popper, and R. R. Fay, Springer, Berlin–Heidelberg–New York, 115–142
- Jekosch, U. (2005) "Assigning meaning to sounds: Semiotics in the context of product- sound design," in: *Communication acoustics*, edited by J. Blauert, Springer, Berlin, Heidelberg, chap. 8, 193–221
- Jones, M. R. & Yee, W. (1993) "Attending to auditory events: The role of temporal organization," in *Thinking in sound: the cognitive psychology of human audition*, edited by S. McAdams and E. Bigand, Clarendon Press, Oxford, chap. 4, .69–112
- Jones, M. R. & Yee, W. (1993), "Attending to auditory events: The role of temporal organization," in: *Thinking in sound: the cognitive psychology of human audition*, edited by S. McAdams and E. Bigand, Clarendon Press, Oxford, chap. 4, 69–112
- Miller, G. A. & Licklider, J.C.R. (1950) "The intelligibility of interrupted speech," *J. Acoust. Soc. Am.* 22, 167–173
- Moore, B., Tyler, L. & Marslen-Wilson, W. (Eds.) (2010) *The perception of speech: From sound to meaning*, Oxford Univ. Press, Oxford
- Nobre, K. & Kastner, S. (2014), *The Oxford handbook of attention*, Oxford Univ. Press, Oxford
- Raake, A. & Blauert, J. (2013) "Comprehensive modeling of the formation process of sound-quality," in *Proc. QoMEX*, Klagenfurt
- Scharf, B. (1970) "Critical bands," in: *Foundations of modern auditory theory*, edited by J. Tobias, Academic Press, New York, vol. 1, 155–202
- Shinn-Cunningham, B. G. (2008) "Object-based auditory and visual attention," *Trends in Cogn. Scs.* 12, 182–186
- Spence, C. & Santangelo, V. (2010) "Auditory attention," in *Oxford handbook of auditory science*, edited by C.J. Plack, Oxford Univ. Press, vol. 3: Hearing, 249–270
- TWO!EARS (2017) "Deliverable D4.3: Final integration-&-evaluation report" URL: <http://twoears.aipa.tu-berlin.de/publications/> – see also URL: <http://twoears.eu>
- Von Ehrenfels, C. (re-edited 1990) *Christian von Ehrenfels: Philosophische Schriften: 4. Metaphysik*, Philosophia Verlag, München
- Warren, R. (1982) *Auditory perception: A new synthesis*, Pergamon Press, New York
- Wolfe, J. M. & Horowitz, T. S. (2004) "What attributes guide the development of visual attention, and how do they do it?" *Nature Rev. Neuroscience* 5, 1–7
- Zwicker, E. (1965) "Temporal effects in simultaneous masking and loudness," *J. Acoust. Soc. Am.* 38, 132–141