

# Nutzung von taktilen semantischen Informationen zur Synthese von plausiblen Ganzkörperschwingungen

Robert Rosenkranz, M. Ercan Altinsoy

*Lehrstuhl für Kommunikationsakustik, TU Dresden, 01062 Dresden, Deutschland,*

*Email: robert.rosenkranz@tu-dresden.de*

## Einleitung

Im letzten Jahrzehnt ließ sich ein beeindruckender Fortschritt von audiovisuellen- sowie Trackingtechnologien beobachten. Dadurch konnten neuartige virtuelle Umgebungen (VU) realisiert werden, die in einer Vielzahl von wichtigen Anwendungsfeldern, wie Fahr-, oder Flugsimulatoren ihren Einsatz finden. Im Gegensatz zur technischen Seite der VU, erfuhren Wahrnehmungsaspekte bisher geringere Aufmerksamkeit. Die inhaltliche Gestaltung von Szenen für die VU, die vom Nutzer stimmig bzw. plausibel wahrgenommen werden ist eine komplexe Aufgabe. Die Erfahrungen und Erwartungen des Nutzers sowie die Interaktionen der einzelnen Modalitäten müssen dabei beachtet werden. Wünschenswert wäre ein perzeptives Synthesemodell, was diese Sachverhalte berücksichtigt. In alltäglichen Situationen, wie bspw. im Straßenfahrzeug, sind Menschen Ganzkörperschwingungen (GKS) ausgesetzt, wodurch bestimmte Wahrnehmungsmerkmale hervorgerufen werden. Diese Wahrnehmungsmerkmale können verbal beschrieben werden, d.h. diese Beschreibungen enthalten semantischen Informationen über die wahrgenommene Situation. Unter Kenntnis der Beziehung zwischen Wahrnehmungsmerkmalen und physikalischen Vibrationssignalen können aus dem Beschreibungprofil einer Szene plausible GKS synthetisiert werden.

## Plausible Synthese

Das Ziel jeder VU ist eine möglichst stimmige Darstellung der Umgebung, so dass Nutzer wie im Alltag auf die Umgebung reagiert bzw. mit dieser Umgebung interagiert. Es gibt zwei Faktoren die dies beeinflussen [1]:

- die „Präsenzillusion“
- die „Plausibilitätsillusion“

Die Präsenzillusion beschreibt den Eindruck in eine Umgebung hineinversetzt zu sein und wird vor allem durch die technische Realisierung der VU begrenzt. Bspw. können Artefakte, die durch eine geringe Bildschirmauflösung oder eine zu große Latenz bei der Reaktion auf Nutzeraktionen entstehen, die Präsenzillusion beeinträchtigen. Die Plausibilitätsillusion bezieht sich auf den Inhalt der dargestellten Umgebung und wird dann beim Nutzer hervorgerufen, wenn die Szene seinen Erwartungen entspricht. Während das Hervorrufen der Präsenzillusion aufgrund der technischen Entwicklung stark vereinfacht wurde, existiert bisher wenig Wissen

darüber, wie man VU gestalten muss, so dass die Erwartungen des Nutzers erfüllt werden.

Es gibt im Wesentlichen zwei Methoden, wie man Szenen gestalten kann [2]:

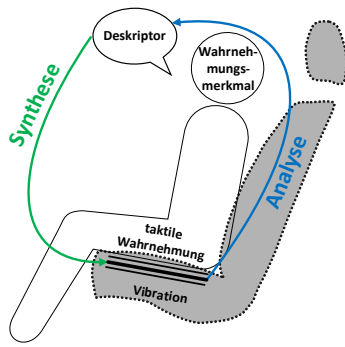
- der „authentische Ansatz“
- der „plausible Ansatz“

Die authentischen Ansatz versucht eine konkrete Szene physikalisch korrekt in der VU zu reproduzieren. Diese Vorgehensweise ist komplex und muss nicht zwangsweise die Erwartungen des Nutzers erfüllen, d.h. eine authentische Darbietung einer Szene wirkt nicht zwangsweise am plausibelsten [3]. Die plausible Ansatz versucht hingegen beim Nutzer die Wahrnehmung hervorzurufen, dass in der VU auftretenden multimodale Ereignisse ihre Entsprechung in einer *vergleichbaren* alltäglichen Umgebung haben. Diese Vorgehensweise hat zwei implizite Vorteile. Erstens müssen lediglich die für die Wahrnehmung des Nutzers relevanten Informationen dargestellt werden und zweitens werden die Erwartungen des Nutzers berücksichtigt.

Es existieren bereits einige Untersuchungen zur plausiblen Synthese von Stimuli. Einen wegweisenden Ansatz für akustische Stimuli stellen die Arbeiten von Gaver dar[4]. Er prägte den Begriff der „Ecological Acoustics“. Dieser besagt, dass in der alltäglichen Umgebung auftretende Geräusche Träger von Informationen über die Umgebung sind, vgl. auch [5]. Aus dieser Beziehung können Ansätze für die plausible Synthese von Geräuschen abgeleitet werden.

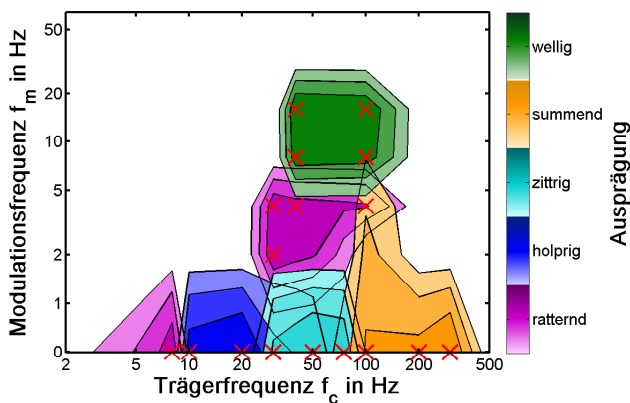
## Wahrnehmungsmerkmale von GKS

In vielen alltäglichen Situationen treten neben Geräuschen auch Vibrationen in Form von Ganzkörperschwingungen (GKS) auf. Diese können ebenfalls Informationen über die Umgebung vermitteln, was in Abb. 1 dargestellt ist: Sobald der Fahrer eines Fahrzeugs das Gaspedal betätigt, werden Vibrationen erzeugt, die vom Fahrer taktil u.a. über den Fahrzeugsitz wahrgenommen werden können. Dadurch werden bei ihm bestimmte Wahrnehmungsmerkmale hervorrufen, welche er in Form von Deskriptoren verbalisieren kann. Die auf ihn einwirkenden Vibrationen vermitteln also semantische Informationen. Der in der vorliegenden Arbeit verfolgte Ansatz versucht die Beziehung zwischen GKS und der durch diese hervorgerufenen Wahrnehmungsmerkmale zur plausiblen Synthese von GKS in VUs einzusetzen.



**Abbildung 1:** In vielen im Alltag auftretenden Situationen werden Menschen GKS ausgesetzt. Dabei stehen die auf den Menschen einwirkenden Vibrationen im Zusammenhang mit den durch diese hervorgerufenen Wahrnehmungsmerkmalen bzw. der verbalen Beschreibung dieser.

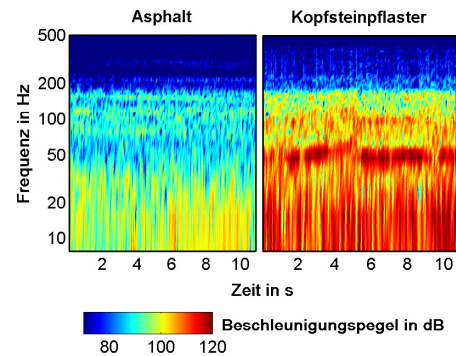
Wahrnehmungsmerkmale von GKS wurden bereits durch Altinsoy [6] untersucht. Es wurden u.a. die sechs häufigsten Wahrnehmungsmerkmale für deterministische periodische Signale und stochastische stationäre Signale ermittelt und in einem Wahrnehmungsexperiment ihre Korrelation mit einfachen Signalmustern überprüft. Der Pegel der präsentierten Signale lag dabei 10 dB über der Wahrnehmungsschwelle. Die Ergebnisse dieser Untersuchung für (amplitudenmodulierte) sinusoidale Signale sind in Abb. 2 dargestellt. Rauschartige Signale wurden etwa zu gleichen Teilen mit den Deskriptoren *ratternd*, *zitterig* und *wummern* beschrieben. Die gefundenen Beziehungen sollen als Ausgangspunkt für die Synthese dienen.



**Abbildung 2:** Beziehung zwischen (amplitudenmodulierten) sinusoidalen Vibrationssignalmustern und den durch sie hervorgerufenen Wahrnehmungsmerkmalen. Die bei der Untersuchung von Altinsoy präsentierten Signale [6] sind durch rote Kreuze repräsentiert. Die mit dem jeweiligen Wahrnehmungsmerkmal korrelierenden Signalbereiche werden durch interpolierte, farbige Bereiche dargestellt, deren Farbintensität abhängig von dem Grad der Korrelation ist.

## Experiment 1 – Bestimmung der Wahrnehmungsmerkmale von GKS im Kontext realer Szenen

Bevor jedoch Szenen synthetisiert werden konnten, mussten zunächst die Wahrnehmungsmerkmale der zu synthetisierenden Szenen in einem Experiment ermittelt werden. Um sicher zu stellen, dass allen Probanden die gleichen Szenen präsentiert werden, mussten Szenen aufgenommen werden. Es wurden zwei gegensätzliche Szenen gewählt: Fahrt eines Mittelklassefahrzeugs auf einer Asphaltstraße bei 80 km/h (Asphaltszene) und auf einer Kopfsteinpflasterstraße bei 30 km/h (Kopfsteinpflasterszene). Für die Aufnahmen wurden ein triaxialer Sitzkissenbeschleunigungssensor auf dem Fahrersitz und zwei Messmikrofone neben den Ohren des Fahrers platziert. Die optische Aufnahme wurde durch eine Kamera realisiert. Die Beschleunigungsspektren der Szenen sind in Abb. 3 zu sehen. Es zeigt sich, dass im wesentlichen rauschartige bzw. sinusoidale Komponenten im Signal enthalten sind.



**Abbildung 3:** Beschleunigungsspektren der aufgenommenen Szenen.

Diese multimodalen Szenen wurden im *Multimodalen Messlabor* des Lehrstuhls für Kommunikationsakustik [7] präsentiert, siehe Abb. 4. Vibrationen von 4 bis 12 Hz wurden vertikal über die Stewart-Plattform und Vibrationen von 12 bis 500 Hz über einen entzerrten elektrodynamischen Shaker dargeboten. Die Wiedergabe von akustischen Stimuli erfolgte als fokussierte Quellen über ein Wellenfeldsynthesesystem und die Wiedergabe von optischen Stimuli über einen HD-Beamer. Die Probanden sollten für jede der zwei Szenen die Ausprägung der Deskriptoren (*holprig*, *summend*, etc.) auf einer kontinuierlichen verbalen Rohrmannskala [8] bewerten. An dem Experiment nahmen sechs Probanden teil.

Die Ergebnisse für die Kopfsteinpflasterszene sind in Abb. 5 zu sehen. Die Probanden fanden die Deskriptoren *holprig*, *ratternd* sowie *wummern* besonders zur Beschreibung der Szene geeignet.

In Abb. 6 sind die Ergebnisse für Asphaltszene dargestellt. Die Deskriptoren mit der stärksten Ausprägung sind *summend* sowie *zitterig*. Im Vergleich mit der Kopfsteinpflasterszene fällt auf, dass die Deskriptorausprägung im Mittel wesentlich geringer ist, was ein Indiz für die Pegelabhängigkeit der Wahrnehmungsmerkmale sein

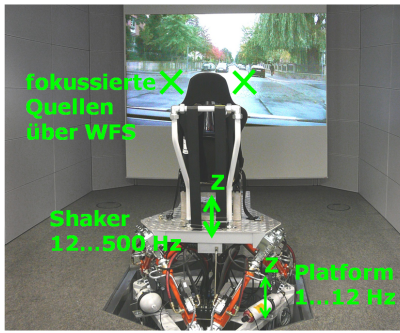


Abbildung 4: Experimentalaufbau im Multimodalen Messlabor, der zur Darbietung der Szenen genutzt wurde.

könnte.

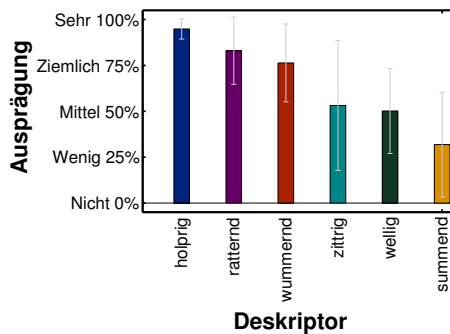


Abbildung 5: Ausprägung der abgefragten Deskriptoren für die Kopfsteinpflasterszene.

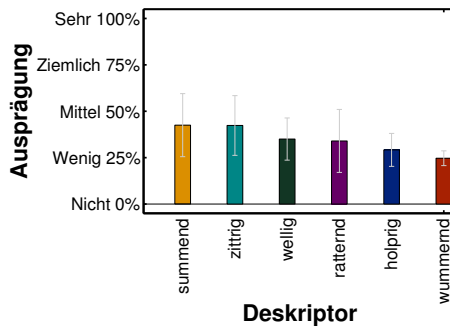


Abbildung 6: Ausprägung der abgefragten Deskriptoren für die Asphaltzene.

## Experiment 2 – Synthese von GKS aus Wahrnehmungsmerkmalen

Die mit den im letzten Abschnitt ermittelten Wahrnehmungsmerkmalen mit der höchsten Ausprägung liefern wahrscheinlich ein Indiz dafür, welche semantischen Informationen dem Nutzer der VU vermittelt werden müssen. Um diese Wahrnehmungsmerkmale hervorzurufen, sollen die mit ihnen am stärksten korrelierenden Signalmuster (siehe Abschnitt Wahrnehmungsmerkmale von GKS) dargeboten werden. Zur Überprüfung dieser Hypothese wurden Vibrationssignale nach einer einfachen Strategie synthetisiert. Zunächst wurde zu jedem Deskriptor ein korrelierendes Signalmuster gewählt, siehe [6]. Da Deskriptoren mit Signalbereichen korrelieren (siehe Abb. 2) wurde jeweils ein Signal aus dem Zentrum des Bereichs gewählt. Da einige Deskriptoren, wie

bspw. ratternd, mit mehreren Signaltypen (sinusoidal, am-sinusoidal, rauschartig) korrelieren, wurde jeweils ein Signal jedes Typs zu diesen Signalen gewählt. Die erzeugten Signale sind in der Übersicht in Abb. 7 zu sehen. Da im vorigen Abschnitt keine Informationen über den Pegel ermittelt wurden, wurden eine einfache Lösung gewählt und die synthetisierten Signale auf den Pegel der Vibrationsaufnahmen der jeweiligen Szenen skaliert.

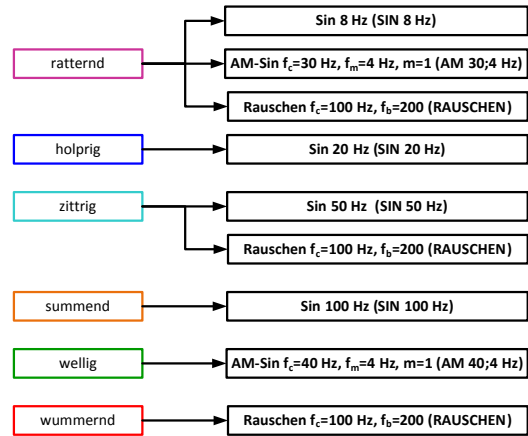


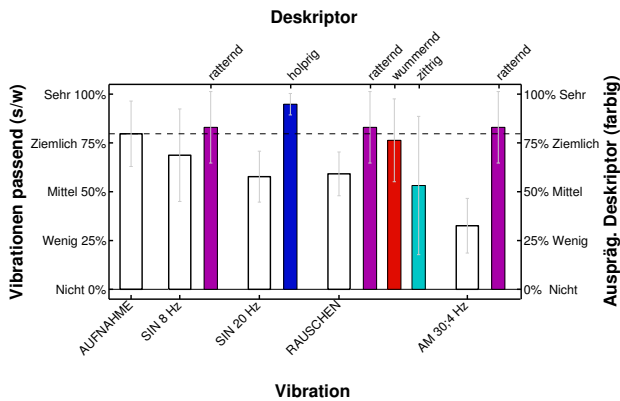
Abbildung 7: Übersicht der aus den Deskriptoren synthetisierten Vibrationen. Für amplitudenmodulierte Signale wurden die Parameter Trägerfrequenz ( $f_c$ ) und Modulationsfrequenz ( $f_m$ ) sowie der Modulationsgrad ( $m$ ) abgekürzt. Für das bandbegrenzte weiße Rauschen wurden die Mittenfrequenz ( $f_c$ ) und die Bandbreite ( $f_b$ ) abgekürzt. In Klammern sind die im Folgenden für die Signale verwendeten Kürzel dargestellt.

Für das folgende Wahrnehmungsexperiment wurde der gleiche Experimentalaufbau wie in Experiment 1 genutzt. Es wurden wieder die Kopfsteinpflasterszene und die Asphaltzene präsentiert. Das Experiment wurde in zwei Sitzungen durchgeführt. In Sitzung 1 wurden den Probanden die multimodalen Szenen mit aufgenommenen Vibrationssignalen dargeboten. In Sitzung 2 wurden den Probanden dieselben Szenen wie in Sitzung 1 präsentiert, jedoch wurden die aufgenommenen Vibrationssignale gegen die synthetisierten Vibrationssignale getauscht. Die Probanden wurden jeweils befragt, wie gut die Vibrationen zur präsentierten Szene passen. Die Bewertung erfolgte ebenfalls auf einer kontinuierlichen verbalen Rohrmannskala [8]. Insgesamt nahmen 6 Probanden an diesem Versuch teil.

Zur Überprüfung des Erfolgs der plausiblen Synthese wurde für jede der zwei Szenen überprüft, ob die mit den Deskriptoren mit der höchsten Ausprägung stark korrelierenden synthetischen Signale am passendsten wahrgenommen wurden. Aus diesem Grund wurden neben den Ergebnissen aus Experiment 1 die Ergebnisse aus Experiment 2 in Abb. 8 bzw. 9 dargestellt. Aus Gründen der Übersichtlichkeit wurden in den Abbildungen die Signale weggelassen, deren korrelierende Deskriptoren nur eine geringe Ausprägung im Verhältnis zu den anderen Deskriptoren besitzen.

Bei der Kopfsteinpflasterszene zeigte sich das das Signal *SIN 8 Hz* fast so passend wie das aufgenommene Signal wahrgenommen wurde. Der mit dem Signal *SIN 8 Hz* korrelierende Deskriptor *ratternd* wurde zur Beschrei-

bung der Szene am 2. geeignetsten wahrgenommen. Die Signale *SIN 20 Hz* und *RAUSCHEN* wurden ähnlich passend wahrgenommen und ihre korrelierenden Deskriptoren haben eine hohe Ausprägung. Im Widerspruch zu Hypothese steht das Signal *AM 30;4 Hz*, dessen korrelierender Deskriptor *ratternd* stark ausgeprägt war. Die Ursache könnte in der Mehrdeutigkeit der Abbildung von dem Deskriptor *ratternd* auf die Signaltypen liegen, siehe Abb. 7. Eine andere Erklärung lieferten die Kommentare der Probanden. Diese nahmen die hochfrequenten synthetischen Signale als sehr regelmäßig bzw. als künstlich und damit weniger passend wahr.

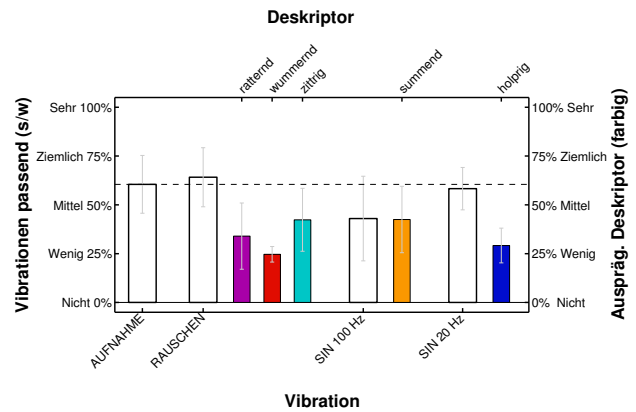


**Abbildung 8:** Eignung der synthetisierten Vibrationssignale der Kopfsteinpflasterszene im Vergleich mit der Ausprägung der korrelierenden Deskriptoren. Die Bewertung der Originalen Vibrationen wurde auf der linken Seite des Diagramms abgetragen.

Bei der Asphaltzene konnte die Hypothese ebenfalls bestätigt werden. Der Deskriptor *zittrig* wurde zur Beschreibung der Asphaltzene am geeignetsten wahrgenommen. Das korrelierende Signalmuster *RAUSCHEN* wurde sogar noch geringfügig besser als das aufgenommene Vibrationssignal bewertet. Das hochfrequente Signal *SIN 100 Hz*, welches mit dem stark ausgeprägten Deskriptor *summend* korreliert, wurde vermutlich ebenfalls als künstlich wahrgenommen und deswegen wenig passend bewertet. Trotz der im Verhältnis zu den Deskriptoren *summend* und *zittrig* geringen Eignung des Deskriptors *holprig* wurde das korrelierende Signal *SIN 20 HZ* als passend wahrgenommen. Möglicherweise ist dies ebenfalls ein Indiz für die Pegelabhängigkeit der Wahrnehmungsmerkmale. Die Bewertung der Szene als wenig *holprig* weist möglicherweise darauf hin, dass das korrelierende Signalmuster mit einem niedrigen Pegel enthalten sein sollte.

## Ausblick

Durch die Experimente konnte gezeigt werden, dass es möglich ist, plausible GKS für VU aus semantischen Information in GKS zu synthetisieren. Die synthetischen Vibrationen wurden ähnlich passend wie die aufgenommenen Vibrationen wahrgenommen. Durch die Berücksichtigung aller Deskriptoren mit hoher Ausprägung, d.h. einer Kombination der korrelierenden Signalmuster, könnte die Synthese noch weiter verbessert werden.



**Abbildung 9:** Eignung der synthetisierten Vibrationssignale der Asphaltzene im Vergleich mit der Ausprägung der korrelierenden Deskriptoren. Anmerkungen siehe Abb. 8.

Die Abbildung von Deskriptoren auf Signalmuster sollte noch erweitert werden, um ein allgemeines Modell zu schaffen. Die gefundenen Zusammenhänge sollten mit einer größeren Anzahl an Szenen verifiziert werden.

## Literatur

- [1] Slater, Mel.: Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments., *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2009, Nr. 1535, S. 3549-3557.
- [2] Novo, P.: *Auditory Virtual Environments*, in Blauert, J. ed., *Communication Acoustics*, Springer Verlag, Berlin, 2005, S. 278
- [3] Rosenkranz, R.; Altinsoy, M.E.; Stamm, M.; Merchel, S.: *Taktile Wahrnehmung im Kontext multimodaler virtueller Umgebungen*, DAGA 2012, Darmstadt, Germany
- [4] Gaver, William W.: *What in the world do we hear?: An ecological approach to auditory event perception*, *Ecological psychology*, 1993, Nr. 1, S. 1-29.
- [5] Jekosch, U.: *Assigning Meaning to Sounds*, in Blauert, J. ed., *Communication Acoustics*, Springer Verlag, Berlin, 2005
- [6] Altinsoy, M.E.: *What can we learn from psychoacoustics regarding the perception of whole-body vibrations? Tactile descriptors for whole-body vibrations*, in *Proceedings of 2nd Polish-German Structured Conference on Acoustics*, 58th OSA, Jurata, Poland.
- [7] Altinsoy, M.E.; Jekosch, U.; Landgraf, J.; Merchel, S.: *Progress in Auditory Perception Research Laboratories — Multimodal Measurement Laboratory of Dresden University of Technology*, 129th AES Convention, San Francisco, CA, USA
- [8] Rohrman, Bernd.: *Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung*, *Zeitschrift für Sozialpsychologie*, 1978, Nr. 3, S. 222-245.