

Über den Zusammenhang zwischen Wahrnehmungsraum und Qualität

Marcel Wältermann, Alexander Raake, Sebastian Möller

Telekom Innovation Laboratories, 10587 Berlin, Deutschland, Email: marcel.waeltermann@tu-berlin.de

Einleitung

Die Wahrnehmung eines physikalischen Ereignisses (hier: eines Schallereignisses) ist im Allgemeinen mehrdimensional. Das Wahrnehmungsereignis lässt sich demnach als ein Punkt im Wahrnehmungsraum darstellen. Aussagen über die Qualität eines Wahrnehmungsereignisses erfordern nach allgemein akzeptierter Vorstellung einen Vergleich zwischen Wahrgenommenem und einer internen Referenz, die die gewünschte Beschaffenheit des Wahrgenommenen beschreibt, also die Erwartung.

Die Trennung zwischen Wahrnehmungsraum und Qualität wirft die Frage nach einer geeigneten Abbildungsfunktion auf. Es ist vorstellbar, dass die Qualität proportional zum Abstand zwischen Wahrnehmungsereignis und einem idealen Punkt im Wahrnehmungsraum abnimmt (Idealpunktmodell). Ein alternatives Modell basiert auf der Annahme eines „Qualitätsvektors“ im Wahrnehmungsraum: Qualität ist hier proportional zu der Projektion eines Punktes im Wahrnehmungsraum auf diesen Vektor (Vektormodell). Diese von J. D. Carroll vereinheitlichten Modellvorstellungen werden in diesem Beitrag rekapituliert. Eine Verallgemeinerung des Idealpunktmodells wird an einem Beispiel aus der Sprachübertragung erläutert.

Versuchspersonenschema

Für das Verständnis der Wahrnehmung eines Schallereignisses soll hier ein systematisches Versuchspersonenschema vorgestellt werden, siehe Abb. 1. Dieses von Blauert [1] vorgeschlagene Schema beschreibt die Wahrnehmungsprozesse einem „Black-Box“-Ansatz folgend. Das räumlich, zeitlich, und eigenschaftlich bestimmte i.A. mehrdimensionale Schallereignis \mathbf{s}_0 durchläuft das „Wahrnehmungssystem“ (W-System) und resultiert gemäß des individuellen Hörvermögens in einem Hörereignis \mathbf{w}_0 . W kann z.B. über psychoakustische Funktionen $\{g_{w,m}\}$ beschrieben werden. Dieses Hörereignis ist i.A. ebenfalls mehrdimensional. Während das Schallereignis durch physikalische Eigenschaften wie den Signal-Rauschabstand oder die spektrale Zusammensetzung beschrieben wird, kann das Hörereignis durch perzeptive Attribute wie etwa die Lautheit oder die Klangfarbe erklärt werden. Sind die Attribute orthogonal, charakterisieren diese die dem Wahrnehmungsraum zu Grunde liegenden Dimensionen: Das Hörereignis kann dann als Punkt im Wahrnehmungsraum dargestellt werden, beschrieben durch den Ortsvektor $\mathbf{w}_0 = [w_{0,1}, \dots, w_{0,m}, \dots, w_{0,M}]^T$. Hier ist M die Anzahl der Dimensionen des zu Grunde liegenden Wahrnehmungsraums. Da das Hörereignis außerhalb des Hörers nicht vorliegt, erhält man Auskunft durch eine Beschrei-

bung β_0 nach Durchlaufen des „Beschreibungssystems“ (B-System) und einer zusätzlichen „Verzerrung“ durch die Skalierung (nicht dargestellt und hier nicht weiter betrachtet; siehe [1] für weitere Details).

Eine Erweiterung des Versuchspersonenschemas wurde von Raake [2] vorgeschlagen, um den Entstehungsprozess eines Qualitätsurteils b eines Schallereignisses abzubilden (z.B. die Qualität übertragener Sprache). Qualität wird hier verstanden als das „Ergebnis der Beurteilung der wahrgenommenen Beschaffenheit einer Einheit im Hinblick auf die erwünschte Beschaffenheit“ (ähnlich [3]).

In Abb. 1 spiegelt das „Vergleichssystem“ (V-System) diese Definition wider, in dem das Hörereignis \mathbf{w}_0 und die Erwartung \mathbf{r}_0 Eingangsgrößen darstellen und im skalaren Qualitätsereignis q_0 resultieren (durch die Versuchsperson beschrieben durch b_0). Die Erwartung \mathbf{r}_0 lässt sich hier, ebenso wie das Hörereignis \mathbf{w}_0 , als Punkt im Wahrnehmungsraum darstellen und durch den Ortsvektor $\mathbf{r}_0 = [r_{0,1}, \dots, r_{0,m}, \dots, r_{0,M}]^T$ beschreiben. Die Erwartung und damit die Qualität sind in hohem Maße kontextabhängig (z.B. von der Anwendung und Situation) und variieren zusätzlich auf Grund persönlicher Einflüsse. Das V-System, das den Wahrnehmungsraum auf die Qualität abbildet, ist hier durch die Funktion g_q charakterisiert:

$$q_0 = g_q(\mathbf{w}_0, \mathbf{r}_0) . \quad (1)$$

Einige Realisierungsmöglichkeiten von g_q werden im Folgenden beschrieben.

Vom Wahrnehmungsraum zur Qualität

J.D. Carroll führt in [4] eine sog. linear-quadratische Hierarchie von Modellen ein, die im Folgenden auszugsweise und in modifizierter Nomenklatur wiedergegeben wird. Insbesondere wird hier auf die Modellierung individueller Differenzen verzichtet und statt dessen ein „Durchschnittshörer“ betrachtet. Eine vollständige Beschreibung der Modellhierarchie findet sich in [4].

Die Modelle zielen auf die Abbildung von Punktkoordinaten $\{w_{0,m}\}$, wie sie bspw. aus der Multidimensionalen Skalierung (MDS) resultieren, auf Qualitätsurteile ab (sog. „External Preference Mapping“). Ausgangspunkt ist die Annahme, dass die Qualität q_0 proportional zum Quadrat der gewichteten Euklidischen Distanz zwischen Hörereignis \mathbf{w}_0 und Erwartung \mathbf{r}_0 ist. Wir verzichten hier auf die Verallgemeinerung der zusätzlichen Rotation des Wahrnehmungsraums [4], womit sich folgendes *Idealpunktmodell* (in [4]: Entfaltungmodell) ergibt:

$$q_0 \sim \sum_{m=1}^M a_m (w_{0,m} - r_{0,m})^2 . \quad (2)$$

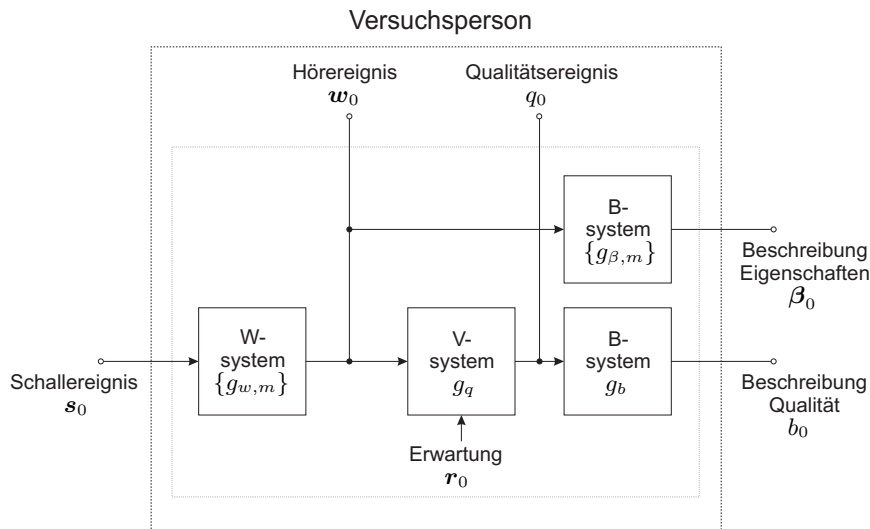


Abbildung 1: Versuchspersonenschema nach [2] (basierend auf [1] und Erweiterungen von [3]).

Gl. (2) spiegelt Jekoschs Qualitätsdefinition [3] in direkter Weise wider: Sowohl die Wahrnehmungskomponenten, als auch die Erwartungskomponenten tauchen im Modell explizit auf, der Vergleich geschieht durch Distanzbestimmung im Wahrnehmungsraum.

Das *Vektormodell* folgt dem Ansatz eines „Qualitätsvektors“ im Wahrnehmungsraum, der in die Richtung optimaler Qualität zeigt. Die Qualität ergibt sich aus dem gewichteten Skalarprodukt $a \langle \mathbf{w}_0, \mathbf{r}_0 \rangle$ aus Hörereignis und Erwartung, d.h. der Projektion des Hörereignisses auf den Qualitätsvektor. Wird die Erwartung für einen gegebenen Kontext als konstant angenommen, ergibt sich das Vektormodell mit $b_m = a \cdot r_{0,m}$ zu:

$$q_0 \sim \sum_{m=1}^M b_m w_{0,m} . \quad (3)$$

Obwohl die Erwartung \mathbf{r}_0 in Form einer Dimensionsgewichtung in das Modell einfließt, ist durch die implizite Referenz ihre Separation von dem Hörereignis, wie es die Definition von Jekosch fordert, nicht gegeben: Die Erwartung ist in diesem Modell unbestimmt. Es ist anzumerken, dass das Vektormodell ein Spezialfall des Idealpunktmodells darstellt. Anschaulich wird dies klar, wenn man sich im Vektormodell einen Idealpunkt weit verschoben in Richtung des Qualitätsvektors vorstellt; die Rangfolge der Distanzen der Hörereignisse zum Idealpunkt nähert sich dann den Projektionen auf den Qualitätsvektor an. Mathematisch wird dies in [4] gezeigt.

Losgelöst von Carrolls Annahmen lässt sich eine Modellgleichung gemäß einer (gewichteten) L_p -Distanz aufstellen, die das Idealpunktmodell bzgl. dessen Metrik verallgemeinert:

$$q_0 \sim \left(\sum_{m=1}^M a_m |w_{0,m} - r_{0,m}|^p \right)^{\frac{1}{p}} . \quad (4)$$

Beispiel: Sprachübertragung

Perzeptive Dimensionen übertragener Sprache können mit „Diskontinuität“, „Rauschhaftigkeit“, und „Klang-

verfärbung“ beschrieben werden und mit kontinuierlichen Skalen mit den Endbezeichnungen *kontinuierlich-diskontinuierlich*, *unrauschhaftig-rauschhaftig* und *klanglich unverfärbt-klanglich verfärbt* in auditiven Experimenten skaliert werden (siehe z.B. [5]). Die skalierten Beschreibungen $\{\beta\}$, erhoben in einem auditiven Experiment mit verschiedenen gestörten Stimuli in einem Breitband-Sprachübertragungskontext, stellen Abbildungen der Hörereignisse $\{\mathbf{w}_0\}$ dar, die allerdings durch die Skalierungs- und Beschreibungsprozesse „verzerrt“ sein können (s.o.).

Mit $r_{0,m} = 0 \forall m$ für diesen Testkontext ergibt das Modell nach Gl. (4) eine hohe Korrelation ($r > 0.95$) zwischen auditiven und prädizierten Werten und $p \approx 1.6$, d.h. eine Euklidische Distanzannahme kann aufrecht erhalten werden [5].

Literatur

- [1] Blauert, J. (1997). *Spatial Hearing — The Psychophysics of Human Sound Localization*. The MIT Press, US-Cambridge MA.
- [2] Raake, A. (2006). *Speech Quality of VoIP — Assessment and Prediction*. John Wiley & Sons, UK-Winchester, West Sussex.
- [3] Jekosch, U. (2005). *Voice and Speech Quality Perception — Assessment and Evaluation*. Signals and Communication Technology. Springer, DE-Berlin.
- [4] Carroll, J. (1972). „Individual Differences and Multidimensional Scaling“. *Multidimensional Scaling — Theory and Applications in the Behavioral Sciences Volume I — Theory*. R.N. Shepard, A.K. Romney, S.B. Nerlove, eds.), 105–155.
- [5] Wältermann, M., Raake, A., Möller, S. (2010). „Analytical Assessment and Distance Modeling of Speech Transmission Quality“. In: *Proc. 11th Annual Conference of the International Speech Communication Association (Inter-speech 2010)*, 1313–1316, JP-Makuhari.