

Vorhersagemodell für Schallquellenexternalisierung in VR/AR-Systemen

Robert Baumgartner¹, Piotr Majdak¹

¹ *Institut für Schallforschung, 1040 Wien, E-Mail: robert.baumgartner@oeaw.ac.at*

Einleitung

Audiowiedergabesysteme im Zusammenhang mit virtueller Realität (VR), erweiterter Realitäten (augmented reality, AR), oder Hear-Through-Kopfhörer verfolgen alle das Ziel Schallquellen möglichst realistisch erscheinen zu lassen. Räumliche Aspekte spielen hier eine zentrale Rolle [1]. Während in natürlichen Hörsituationen Schallquellen üblicherweise externalisiert wahrgenommen werden, d.h. einer räumlichen Position außerhalb des Kopfes zugeordnet werden können, resultiert Kopfhörerwiedergabe oft in einer Im-Kopf-Lokalisation. Eine umfangreiche Evaluierung obiger Wiedergabesysteme gestaltet sich besonders aufwändig, wenn Hörversuche erforderlich werden. Vorhersagemodelle können hier den Entwicklungsprozess sehr effizient unterstützen. Da spektrale Merkmale aufgrund ihrer Individualität oft besonders problematisch zu reproduzieren sind, skizzieren wir hier ein phänomenologisches Modell zur Prognose von Externalisierungsdefiziten mit Fokus auf spektralen Merkmalen und unter der Annahme einer statischen und somit besonders kritischen Hörsituation.

Die externalisierte Wahrnehmung von Schallquellen wurde bisher zumeist als auditorische Distanz zwischen unterschiedlich genauen virtuellen Repräsentationen einer tatsächlichen Quelle erfasst [2]–[7]. Diese Studien zeigen unter anderem, dass externalisierte Wahrnehmung robust ist gegenüber spektralen Veränderungen der interauralen Zeitdifferenzen (interaural time differences, ITDs), jedoch stark durch spektrale Modifikationen in interauralen Pegeldifferenzen (interaural level differences, ILDs) und monauralen Amplitudenspektren beeinträchtigt werden kann.

Während bis dato nur die Verarbeitung von ILD-Spektren in einem Externalisierungsmodell Berücksichtigung fand [4], sollen hier auch die monauralen spektralen Merkmale berücksichtigt werden. Vergangene Modellstudien zum Richtungshören haben die Verarbeitung der monauralen spektralen Merkmale speziell für die Schallquellenlokalisation entlang der vertikalen Dimension untersucht [8]–[10]. Sowohl Tierversuche [11] als auch Modellsimulationen [8] legen nahe, dass neuronale Prozesse auf der Analyse von positiven spektralen Gradienten basieren. Wir übernehmen hier dieses monaurale Verarbeitungsprinzip obwohl es bisher noch keine klare Evidenz dafür gibt, dass diese Ergebnisse auf das menschliche auditorische System übertragen werden können und inwiefern diese Merkmale auch von Bedeutung für Externalisierungswahrnehmung sind. Die hier berichteten Simulationsergebnisse sind ein erstes Indiz für die Generalisierbarkeit.

Das Modell

Die Modellstruktur folgt in ihren Grundzügen dem Prinzip des Mustervorlagenvergleichs (template matching): Merkmale des eintreffenden Signals werden mit Mustervorlagen für die entsprechenden Merkmale verglichen und dabei eine Abweichungsgröße generiert (siehe Abb. 1). Als Merkmale werden die positiven Gradienten der Amplitudenspektren und die spektralen ILDs verarbeitet. Die Abweichungsgrößen werden bei der monauralen Verarbeitung getrennt für jedes Ohr berechnet und dann mit einer binauralen Gewichtung, die den Beitrag des ipsilateralen Ohrs mit zunehmendem Lateralwinkel verstärkt, zusammen addiert. Die resultierende Abweichungsgröße wird anschließend mit einer exponentiell abfallenden Funktion auf das Externalisierungsmaß abgebildet. Die Steilheit der Abbildungsfunktion wird durch einen Sensitivitätsparameter bestimmt, der individuell für jedes Merkmal angepasst wurde.

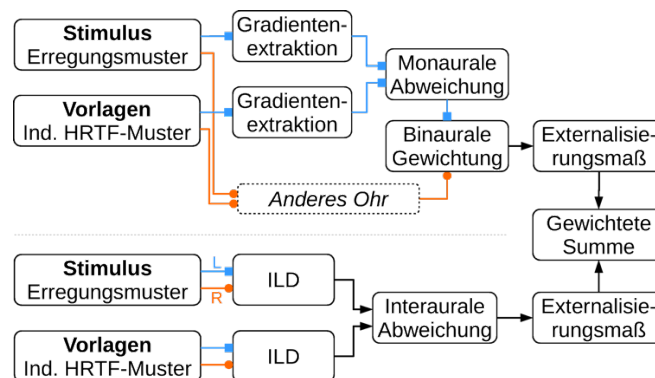


Abbildung 1: Modellstruktur zur Verarbeitung monauraler (oben) und interauraler (unten) spektraler Merkmale.

Für die Anpassung der Sensitivitätsparameter wurde ein Vorhersagefehler ermittelt, der dem Effektivwert der Differenzen zwischen den tatsächlichen und prognostizierten Externalisierungsmaßen entspricht. Die optimalen Sensitivitätsparameter wurden durch Minimierung der quadratischen Vorhersagefehler gewonnen.

Die aus den beiden Merkmalen resultierenden Externalisierungsmaße wurden anschließend gewichtet addiert, wobei die Gewichte wiederum so gewählt wurden, dass sich ein minimaler quadratischer Vorhersagefehler ergab.

Zur Modellevaluierung wurden insgesamt fünf Experimente aus vier Studien herangezogen, welche sich alle mit der Auswirkung von spektralen Modifikationen (wie spektraler Glättung, Färbung und Tiefpassfilterung) auseinandersetzen und unter statischen Bedingungen sowie konstantem Nachhall durchgeführt wurden [2]–[5].

Resultate

Die Evaluierungsergebnisse haben gezeigt, dass das Modell nur dann gut an alle Daten angepasst werden konnte, wenn beide Merkmale berücksichtigt wurden. Damit konnte ein mittlerer Vorhersagefehler von nur 11% erzielt werden, wenn die Externalisierungsprognosen basierend auf monauralen Merkmalen mit etwa 60% und jene von interauralen Merkmalen mit etwa 40% gewichtet wurden. Der mittlere Fehler erhöhte sich auf 14% Prozent, wenn nur monaurale Merkmale Verwendung fanden, und auf 18%, wenn nur interaurale Merkmale berücksichtigt wurden. Gleichzeitig verdoppelte sich der Standardfehler von etwa 5% auf 10% über die Experimente hinweg, was rückschließen lässt, dass die einzelnen Experimente die beiden Merkmale sehr unterschiedlich stark beeinflussten.

Diskussion und Ausblick

Im Gegensatz zu bisher existierenden Modellen zur Externalisierungswahrnehmung wurden hier sowohl interaurale als auch monaurale spektrale Merkmale ausgewertet und beide erwiesen sich als erforderlich um die Vielzahl an hier inkludierten experimentellen Daten erklären zu können.

Die Implementierung zur Auswertung monauraler spektraler Merkmale wurde von einem Modell für Schallquellenlokalisation in Sagittalebene übernommen und erschien, den guten Ergebnissen zufolge, als generalisierbar auf die Externalisierungswahrnehmung. Um jedoch die Implementierung hinsichtlich der Fokussierung auf positiven spektralen Gradienten noch genauer zu untersuchen, werden derzeit Experimente durchgeführt. Weiters werden momentan Erweiterungen des Modells vorangetrieben, welche auch Unterschiede im Nachhall berücksichtigen.

Literatur

- [1] P. Majdak, R. Baumgartner, and C. Jenny, "Formation of three-dimensional auditory space," in *The Technology of Binaural Understanding*, Springer, Berlin, Heidelberg, 2020.
- [2] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J Acoust Soc Am*, vol. 99, no. 6, pp. 3678–88, Jun. 1996.
- [3] A.W. Boyd, W. M. Whitmer, J. J. Soraghan, and M. A. Akeroyd, "Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers," *J. Acoust. Soc. Am.*, vol. 131, no. 3, pp. EL268–EL274, Mar. 2012, doi: 10.1121/1.3687015.
- [4] H. G. Hassager, F. Gran, and T. Dau, "The role of spectral detail in the binaural transfer function on perceived externalization in a reverberant environment," *J. Acoust. Soc. Am.*, vol. 139, no. 5, pp. 2992–3000, May 2016, doi: 10.1121/1.4950847.
- [5] R. Baumgartner et al., "Asymmetries in behavioral and neural responses to spectral cues demonstrate the generality of auditory looming bias," *Proc. Natl. Acad. Sci.*, vol. 114, no. 36, pp. 9743–9748, May 2017, doi: 10.1073/pnas.1703247114.
- [6] J. Catic, S. Santurette, J. M. Buchholz, F. Gran, and T. Dau, "The effect of interaural-level-difference fluctuations on the externalization of sound," *J. Acoust. Soc. Am.*, vol. 134, no. 2, pp. 1232–1241, Aug. 2013, doi: 10.1121/1.4812264.
- [7] J. C. Gil-Carvajal, J. Cubick, S. Santurette, and T. Dau, "Spatial Hearing with Incongruent Visual or Auditory Room Cues," *Sci. Rep.*, vol. 6, p. 37342, Nov. 2016, doi: 10.1038/srep37342.
- [8] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *J. Acoust. Soc. Am.*, vol. 136, no. 2, pp. 791–802, Aug. 2014, doi: 10.1121/1.4887447.
- [9] R. Baumgartner, P. Majdak, and B. Laback, "Modeling the Effects of Sensorineural Hearing Loss on Sound Localization in the Median Plane," *Trends Hear.*, vol. 20, pp. 1–11, 2016, doi: 10.1177/2331216516662003.
- [10] R. Baumgartner, P. Majdak, and B. Laback, "Erratum: Modeling sound-source localization in sagittal planes for human listeners [*J. Acoust. Soc. Am.* 136, 791–802 (2014)]," *J. Acoust. Soc. Am.*, vol. 140, no. 4, pp. 2456–2456, Oct. 2016, doi: 10.1121/1.4964753.
- [11] L. A. J. Reiss and E. D. Young, "Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus," *J. Neurosci.*, vol. 25, no. 14, pp. 3680–3691, Feb. 2005, doi: 10.1523/JNEUROSCI.4963-04.2005.