

# Bewegungsprädiktion in der Echtzeit-Auralisierung dynamischer Schallfelder

Frank Wefers, Michael Vorländer  
 Institut für Technische Akustik, RWTH Aachen,  
 Kopernikusstr. 5, 52074 Aachen, Deutschland,  
 Email: fwe@akustik.rwth-aachen.de

## Einleitung

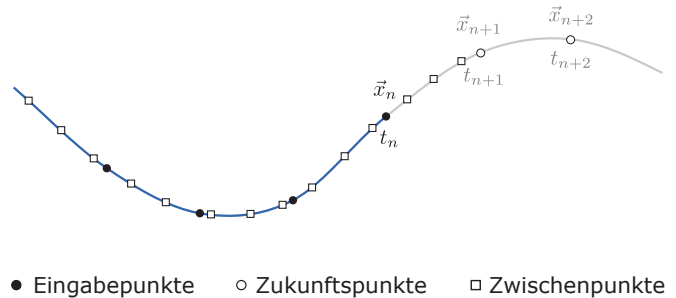
Die Auralisierung bewegter Schallquellen und Hörer erfordert Kenntnis der effektiven Ausbreitungsdauer  $\tau(t)$  der Schallwellen im Medium (*retardierte Zeit*). Physikalisch konsistente Schallfeldsimulation können erreicht werden indem man letztere mittels variabler Laufzeitglieder (*variable delay-lines*) realisiert. Allerdings erfordert dies die Kenntnis der Ausbreitungsdauer in jedem Abtastwert (*sample*) des Audiosignals. Im Fall von Offline-Simulationen liegen abgetastete Bewegungs-Trajektorien der Objekte (Quellen, Hörer) bereits vollständig vor. Diese können mit Hilfe des Abtasttheorems (Tiefpass-Rekonstruktion) in kontinuierliche Bewegungs-beschreibungen überführt werden. Ausgehend von dieser Basis kann  $\tau(t)$  als Lösung nicht-linearer Gleichungen (iterativ) bestimmt werden [7].

Betrachtet man hingegen Echtzeit-Systeme, so sind die Trajektorien zwar ebenfalls abgetastet, aber nur bis zum gegenwärtigen Zeitpunkt bekannt. Auch hier wird eine kontinuierliche Bewegungsbeschreibung benötigt, um beliebige Zeitpunkte in anderen Zeitskalen abzutasten (Bewegungs-Interpolation). Um die Latenz in virtuellen Umgebungen gering zu halten, können Bewegungsprädiktion angewendet werden (e.g. *predictive tracking for visual rendering*, Übersprechkompensation). Die hier betrachtete Anwendung unterscheidet sich in den Anforderungen (e.g. maximale Konvergenz vs. Stetigkeit in höheren Momenten). Der Beitrag untersucht die Eignung *exponential smoothing* für solche Zwecke der Schallfeldsimulation und die Berechnung von  $\tau(t)$ .

## Bewegung von Quellen und Hörern

Objekte in dynamischen Szenen ändern mit der Zeit ihre Position und Lage im Raum. Eine Trajektorie (Bahnkurve) beschreibt hierbei die Position  $\vec{r}(t) \in \mathbb{R}^3$  in Abhängigkeit der Zeit  $t \in \mathbb{R}^+$ . Die Orientierung des Objektes wird hier nicht näher betrachtet. Im Folgenden wird der vereinfachte Fall einer einzelnen Quelle (*source*, S) und eines einzelnen Hörers bzw. Empfängers (*receiver*, R) betrachtet. Die Trajektorien beider Objekte seien als  $\vec{r}_S(t)$  und  $\vec{r}_R(t)$  bezeichnet. Alle weiteren Überlegungen gelten für Freifeldbedingungen (keine Reflexionen) und ein ruhendes, homogenes fluides Medium mit einer konstanten Schallgeschwindigkeit  $c$ . Die Geschwindigkeiten  $\vec{v}_S(t)$ ,  $\vec{v}_R(t)$  der Objekte seien bezüglich dieser signifikant kleiner  $\|\vec{v}_S(t)\|, \|\vec{v}_R(t)\| \ll c$ .

Strauss [7] befasst sich mit der Simulation solcher Umgebungen. Ein Schallereignis, welches der Hörer zur Zeit



• Eingabepunkte    ◦ Zukunftspunkte    ◻ Zwischenpunkte

Abbildung 1: Abgetastete Trajektorie eines Objektes.

$t$  am Ort  $\vec{r}_R(t)$  wahrnimmt, kann einem eindeutigen Ursprung zugeordnet werden. Angenommen, die betreffende Wellenfront unterlag einer Ausbreitungsdauer  $\tau(t)$ , so liegt der Ursprung der Welle in der Zeit  $t - \tau(t)$  am Ort  $\vec{r}_S(t - \tau(t))$ . Eine physikalisch korrekte Auralisierung zwischen dem Zeitpunkten der Schallabstrahlung  $t - \tau(t)$ , der Ausbreitungsphase  $\tau(t)$  und dem Zeitpunkt des Eintreffens beim Hörer  $t$  unterscheiden. Echtzeitfähige Datenstrukturen für solche interaktiven Schallfeldsimulationen werden in [8] vorgestellt.

Sind die Trajektorien der Objekte a-priori bekannt (z.B. bei Offline-Auralisierung), so kann die Ausbreitungsdauer als Lösung der folgenden nicht-linearen Gleichung bestimmt werden [7]

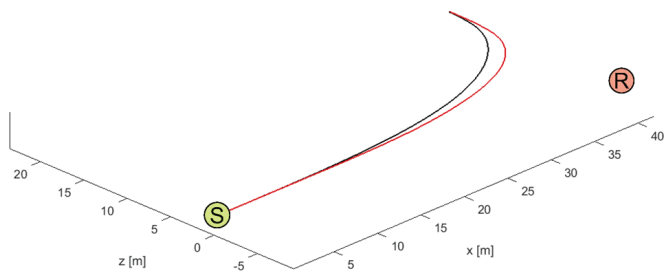
$$\|\vec{r}_R(t) - \vec{r}_S(t - \tau(t))\| = c\tau(t) \quad (1)$$

Hierzu muss ein mathematisches Modell für die Trajektorien angenommen werden. In [7] werden hierfür lineare gleichförmige Bewegungen mit konstanter Geschwindigkeit (*constant velocities, CV*) der Form

$$\vec{r}(t) = \vec{x} + \vec{v}(t) \quad (2)$$

angesetzt. Ist die Trajektorie abgetastet, so kann die Lösung für  $\tau(t)$  auf Basis der Lösungen für aller Abschnitte bestimmt werden. In der Signalverarbeitungskette wird  $\tau(t)$  durch ein variables Laufzeitglied (*variable delay-line, VDL*) realisiert. Hierzu muss  $\tau(t)$  mit der Zeitaufösung der Blöcke des Audiodatenstroms (*frames*) oder sogar einzelner Abtastwerte (*samples*) bestimmt werden. Die zeitabhängige Ausbreitungsdauer  $\tau(t)$  beeinflusst die vom Hörer wahrgenommenen Frequenzen (Dopplereffekt). Der Verhältnis zwischen einer von der Quelle abgestrahlter Frequenz und der vom Hörer wahrgenommenen berechnet sich über die zeitliche Ableitung der Ausbreitungsdauer [7]

$$\frac{f_R}{f_S}(t) = 1 - \frac{d\tau(t)}{dt} \quad (3)$$



**Abbildung 2:** Trajektorie eines Kraftfahrzeugs (S) in Kurvenfahrt (schwarz). Der Hörer ruht an der Position (R). Rot dargestellt die Echtzeit-Prädiktion mittels Double Exponential smoothing (DES).

In Echtzeitsystemen ergeben sich völlig neue Problemstellungen, welche in Abbildung 1 illustriert sind: Der zukünftige Verlauf (o) der Bahnkurven ist nicht bekannt, sondern wurde nur zu vergangenen Zeiten eingegeben (•). Trotzdem verlangt es die Echtzeit-Auralisierung die Ausbreitungsdauer fortlaufend zu berechnen. Daher müssen Zwischenpunkten der Trajektorie berechnet werden (□). Dieser Beitrag untersucht wie dies unter Echtzeitbedingungen geschehen kann und geht die auf physikalischen und perceptiven Zusammenhänge ein (z.B. Stetigkeit, Frequenzänderungen).

## Fallbeispiel

Zur näheren Untersuchung wurde die Trajektorie eines Kraftfahrzeuges gewählt (Abbildung 2, schwarze Kurve). Sie bildet eine Abbiegesituation nach (Bremsen, Einlenken, Beschleunigen). Der Hörer befindet sich an einer konstanten Position und hört den Schall des vorbeifahrenden Fahrzeugs. Basierend auf echten Fahrzeugdaten (Längst- und Querschleunigungen) wurde die Trajektorie analytisch definiert und kann somit in beliebiger Zeitaufösung ausgewertet werden (Referenz). Ihre physikalischen Daten sind Abbildung 4 zu entnehmen. Zur perceptiven Bewertung wurde eine stark vereinfachte Auralisierung durchgeführt. Hierzu wurde für die Quelle (S) durch einen 730 Hz Sinuston modelliert und nur die Frequenzmodulation (Gleichung 3) und die Pegelabnahme nach dem  $1/r$  Abstandsgesetz simuliert. Mehrere Personen führten subjektive Vergleiche der generierten Signale durch und bewerteten diese hinsichtlich Artefakten. Umfassende Hörversuche wurden noch nicht durchgeführt und sind für die Zukunft geplant.

## Interpolation von Trajektorien

Zunächst wird der Fall der Zwischenwertberechnung auf Trajektorien behandelt. Dieser findet auch in den nachfolgend behandelten Prädiktionsverfahren Anwendung. Ein Interpolationsproblem ist wie folgt definiert: Die Trajektorie ist gegeben durch eine Folge von Abtastwerten  $(t_n, \vec{x}_n)$ . Zu einer vorgegebenen Zeit  $t$ , welche in das Segment  $t_n \leq t < t_{n+1}$  fällt, soll nun die Zwischenposition  $\vec{r}(t)$  berechnet werden. Der einfachste Ansatz ist eine lineare Interpolation, gegeben durch

$$\vec{r}(t) = \vec{x}_n + \vec{v}_n(t - t_n) \quad \text{mit} \quad \vec{v}_n = \frac{\vec{x}_{n+1} - \vec{x}_n}{t_{n+1} - t_n} \quad (4)$$

Die Tangente  $\vec{v}_n$  (Differenzenquotient) entspricht hierbei der der Geschwindigkeit  $\partial \vec{r}(t)/\partial t$ , welche im Segment  $[t_n, t_{n+1}]$  konstant ist. Anders ausgedrückt nimmt die lineare Interpolation eine gleichförmige Bewegung mit konstanter Geschwindigkeit an. Abbildung 3a verdeutlicht dies anhand der eingeführten Fahrzeugtrajektorie. Die Geschwindigkeiten  $\vec{v}_n$  werden unabhängig für jedes Segment berechnet und erfüllen damit keine Stetigkeits- oder Übergangskriterien. Mathematisch ausgedrückt ist die lineare Interpolation nur  $C^0$ -stetig und verläuft als geschlossener Geradenzug durch die vorgegebenen Kontrollpunkte. Die Geschwindigkeit (und damit auch die Beschleunigung) kann sich sprunghaft von Segment zu Segment ändern (Abbildung 3a). Als Konsequenz aus Gleichung 3 treten damit auch Sprünge in den wahrgenommenen Frequenzen auf (Abbildung 3b). Die durchgeführten Hörtest (siehe vorheriger Abschnitt) zeigten, dass diese hörbar sein können. Jedoch ist noch nicht näher untersucht worden, wie dies im Zusammenhang steht mit der Abtastrate der Trajektorie steht. Höhere Abtastraten mit feinerem Zeitraster verringern diese Effekte, so dass diese unter die Wahrnehmungsschwelle fallen können.

Die Suche nach besser geeigneten Interpolationsverfahren führte zur Catmull-Rom Splines [3, 5], im Folgenden abgekürzt als CR. Dieses stammen ursprünglich aus der Computergrafik und werden zur Berechnung von Zwischenpunkten in der Animation verwendet (sogenanntes *in-betweening*) [1]. Sie basieren auf kubischen Hermiteschen Splines, Polynomen dritten Grades, welche nicht nur durch die Kontrollpunkte  $(t_n, \vec{x}_n)$  verlaufen, sondern dort auch vorgegebene Ableitungen (in diesem Kontext Geschwindigkeitsvektoren)  $(t_n, \vec{v}_n)$  einhalten. Auf die Berechnung wird hier nur oberflächlich eingegangen. Details finden sich in [5]. Die hier verwendete Nomenklatur unterscheidet sich und ist an die physikalischen Begriffe angelehnt. Ausgehend von einer normierten Zeit  $\tilde{t}$

$$\tilde{t} = \frac{t - t_n}{t_{n+1} - t_n} \in [0, 1] \quad (5)$$

berechnen sich die Zwischenwerte als Wert des Polynoms

$$\vec{r}(t) = \vec{c}_0 + \vec{c}_1 \tilde{t} + \vec{c}_2 \tilde{t}^2 + \vec{c}_3 \tilde{t}^3 \quad (6)$$

Die Koeffizienten  $\vec{c}_i$  berechnen sich wie folgt (hier vereinfacht dargestellt für eine einzelne Koordinatenachse)

$$\begin{bmatrix} c_3 \\ c_2 \\ c_1 \\ c_0 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 1 & 1 \\ -3 & 3 & -2 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} x_n \\ x_{n+1} \\ v_n \\ v_{n+1} \end{bmatrix} \quad (7)$$

Catmull-Rom Splines bestimmen die Geschwindigkeiten an den Segmentgrenzen mittels finiter Zentralfdifferenzen

$$\vec{v}_n = \frac{\vec{x}_{n+1} - \vec{x}_{n-1}}{t_{n+1} - t_{n-1}}, \quad \vec{v}_{n+1} = \frac{\vec{x}_{n+2} - \vec{x}_n}{t_{n+2} - t_n} \quad (8)$$

Sie sind  $C^1$ -stetig. Sowohl Positionen, als auch Geschwindigkeiten sind geschlossene Kurven. Es treten keine Sprünge in den Geschwindigkeiten auf. Innerhalb eines Segmentes hat die Beschleunigung  $\partial \vec{v}_n / \partial t$  die lineare

Form  $\vec{a}_{0,n} + \vec{a}_n t$ . Abbildung 4 zeigt die Ergebnisse der CR-Interpolation der Fahrzeugtrajektorie. Es ist ersichtlich, dass sowohl die Geschwindigkeit, als auch das resultierende Frequenzverhältnis glatt verlaufen und keine Unstetigkeiten aufweisen. Probanden konnten in Hörtests keine Unterschiede zur analytischen Lösung feststellen. Damit bleibt festzuhalten, dass die CR-Interpolation hinsichtlich der physikalischen und perceptiven Merkmale ein geeignetes Interpolationsverfahren für die Auralisierung darstellt.

Prinzipiell lässt sich auf diese Weise eine robuste Auralisierung bewegter Quellen konstruieren. Allerdings setzt Interpolation die Kenntnis zukünftiger Kontrollpunkte voraus (siehe Abbildung 1). Lineare Interpolation erfordert dies die Kenntnis zukünftiger Werte. Bei CR-Interpolation sind es zwei Zukunftswerte (aufgrund der zentralen Differenzenquotienten in Gleichung 8). Nimmt man beispielsweise einen Motion-Tracker mit 30 Hz Abtastrate an, so geht dieser Ansatz mit einer zusätzlichen Latenz von  $2/30 \text{ Hz} \approx 66 \text{ ms}$  einher. Dies kann durchaus im Bereich der Wahrnehmbarkeit liegen und eine audiovisuelle Darbietung degradieren (*audio-visual mismatch*).

### Prädiktion von Trajektorien

Eine alternative Vorgehensweise ist die benötigten Zukunftswerte entsprechend vorherzusagen (Prädiktion). In der Virtuellen Realität (VR) hat sich die Vorhersage mittels doppelter exponentieller Gewichtung (*double exponential smoothing prediction, DESP*) [4] als geeignetes Mittel zur Positionsprädiktion etabliert [6, 2]. Dessen Eignung zur Echtzeitprädiktion von Trajektorien für die Auralisierung wurde hier untersucht. Das Prädiktionsverfahren arbeitet wie folgt: Aus der Sequenz der eingegebenen Positionen  $\vec{x}_0, \vec{x}_1, \dots$  werden zwei gefilterte Sequenzen  $\vec{S}_n$  und  $\vec{S}_n^{(2)}$  definiert [6]

$$\vec{S}_n = \alpha \cdot \vec{x}_n + (1 - \alpha) \cdot \vec{S}_{n-1} \quad (9)$$

$$\vec{S}_n^{(2)} = \alpha \cdot \vec{S}_n + (1 - \alpha) \cdot \vec{S}_{n-1}^{(2)} \quad (10)$$

Die für den Zeitpunkt  $t_{n+1}$  vorhergesagte Position  $\vec{x}_{n+1}$  berechnet sich nach [6]

$$\vec{x}_{n+1} = \left[ 2\vec{S}_t - \vec{S}_t^{(2)} \right] + \left[ \frac{\alpha}{1 - \alpha} (\vec{S}_t - \vec{S}_t^{(2)}) \right] \quad (11)$$

Der Parameter  $\alpha \in [0, 1]$  (*smoothing factor*) bestimmt wie stark vergangene Werte gewichtet werden. Kleine Werte von  $\alpha$  führen zu starker Einbeziehung vergangener Werte und einer robusten Vorhersage. Allerdings reagiert die Prädiktion in diesem Fall träge auf spontane Änderungen. Große Werte von  $\alpha$  konzentrieren die Vorhersage auf zeitnahe Werte. Sie reduzieren die Reaktionszeit des Prädiktors, machen aber die Vorhersage unsicherer. Beiden Wertefolgen  $\vec{S}_n, \vec{S}_n^{(2)}$  ist kein direkter physikalischer Zusammenhang zugeordnet.

Allgemein hängt die geeignete Wahl des Parameters  $\alpha$  von der Charakteristik der Daten ab. Tests mit kleinen Werten ( $\alpha < 0.5$ ) zeigten sich ungeeignet für die

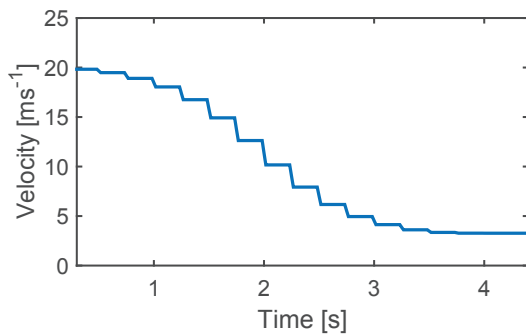
betrachtete Fahrzeugtrajektorie und führt zu signifikanten Prädiktionsfehlern. Abweichungen von einigen Metern und deutlich unterschiedliche Geschwindigkeiten waren die Konsequenz. Hieraus ergeben sich Folgefehler, wie z.B. falsche relative Richtungen der Quelle aus Sicht des Hörers, Einflüsse auf die wahrgenommene Lautstärke ( $1/r$  Gesetz) und veränderte Frequenzverhältnisse beim Dopplereffekt. Indes stellt die Trägheit des Schätzer (Gruppenlaufzeit aus systemtheoretischer Sicht) die ursprüngliche Motivation (Latenzvermeidung) in Frage.

Brauchbare Ergebnisse wurden für  $\alpha = 0.9$  gefunden. Die prädizierte Trajektorie ist als rote Kurve in Abbildung 1 gezeigt. Vor dem Einlenken bremst das Fahrzeug ab. Deutlich zu sehen ist die verspätete Reaktion des Prädiktors (die rote Kurve knickt später ein). Dies ist auch erkennbar im prädizierten Verlauf der Geschwindigkeit (Abbildung 5a), welches andere Frequenzverschiebungen zur Folge hat (Abbildung 5b). Die durch den Prädiktor verursachte Verzögerung wurde mit ca. 150 ms bestimmt. Dies ist deutlich größer als der oben bezifferte Wert für die Strategie mittels Interpolation. Das Unterschwingen des Frequenzverhältnisses führte zu einer Abweichung von -10 Cent bis +20 Cent (musikalisches Intervall) vom korrekten Verhältnis. Alle Versuchspersonen waren in der Lage die Auralisierung der prädizierten Trajektorie von der tatsächlichen zu unterscheiden.

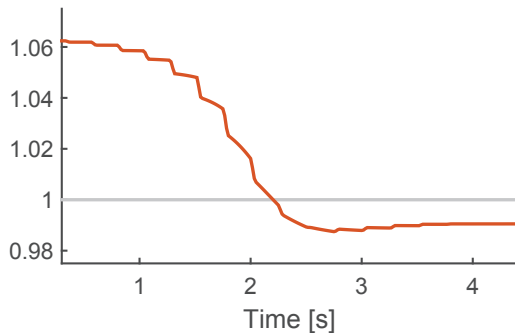
### Zusammenfassung

Untersucht wurde die Beschreibung, Interpolation und Prädiktion von Trajektorien zur Echtzeit-Auralisierung bewegter Quellen und Hörer. Die notwendige Zwischenwertberechnung muss physikalischer Zusammenhänge und daraus resultierende perzeptive Parametern berücksichtigen. Lineare Interpolation führt zu Unstetigkeiten der Geschwindigkeit und nimmt somit Einfluss auf der Verhältnis der Quelle abstrahlten und vom Hörer wahrgenommenen Frequenzen (Dopplereffekt). Diese Artefakte können hörbar sein. Interpolation mittels Catmull-Rom Splines ist aus akustischer Sicht besser geeignet, da sie Stetigkeit in Positionen und Geschwindigkeiten garantiert. Das Frequenzverhältnis behält damit einen glatten zeitlichen Verlauf.

Interpolation von Trajektorien kann auch in Echtzeitsystemen eingesetzt werden, soweit die resultierende Verzögerung der Auralisierung tolerierbar ist. Alternativ können Prädiktionsverfahren verwendet werden. Im hier betrachteten Beispiel stellte sich dies aber als nicht-trivial heraus. Eine ausreichend robuste Vorhersage der Trajektorie mittels *double exponential smoothing* konnte nur mittels eines vergleichsweise trägen Prädiktors erreicht werden, dessen Verzögerung oberhalb der Latenz des vergleichbaren Interpolationsansatzes lag. Welche Latenzen speziell für den Fall der Auralisierung bewegter Quellen und Hörer akzeptabel sind, muss in Zukunft noch näher untersucht werden. Es ist zu klären ob geeignetere Prädiktoren existieren, welche ausreichend akkurate Vorhersagen erlauben und so schnell reagieren, dass sie eine echte Alternative zum Interpolationsansatz darstellen.



(a) Relativgeschwindigkeit  $\|\vec{v}_{SR}\|$

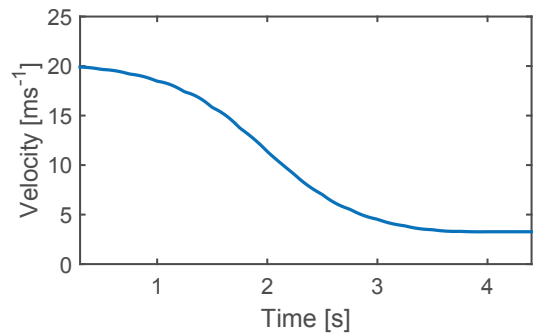


(b) Doppler-Frequenzverhältnis  $f_R/f_S$

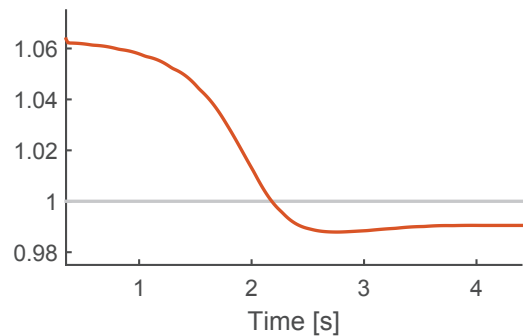
**Abbildung 3:** Physikalische Größen für die lineare Interpolation der Trajektorie in Abbildung 2.

**Literatur**

- [1] Watt Alan and Watt Mark. Advanced animation and rendering techniques. Addison-Wesley, 1992.
- [2] Ingo Assenmacher. Low latency technology for interactive virtual environments. PhD thesis, Doktorarbeit, RWTH Aachen, 2009.
- [3] E. Catmull and R. Rom. A class of local interpolating splines. Computer aided geometric design, 74:317–326, 1974.
- [4] E. S. Gardner and D. G. Dannenbring. Forecasting with exponential smoothing: Some guidelines for model selection. Decision Sciences, 11(2):370–383, 1980.
- [5] D. Kochanek and R. Bartels. Interpolating splines with local tension, continuity, and bias control. In ACM SIGGRAPH Computer Graphics, volume 18, pages 33–41. ACM, 1984.
- [6] J. LaViola. Double exponential smoothing: an alternative to kalman filter-based predictive tracking. In Proceedings of the workshop on Virtual environments 2003, pages 199–206. ACM, 2003.
- [7] H. Strauss. Implementing Doppler Shifts for Virtual Auditory Environments. In Audio Engineering Society Convention 104, 1998.
- [8] F. Wefers and M. Vorländer. Zeitvariante Beschreibung virtueller Szenen für die Echtzeit-Auralisierung instationärer Schallfelder. In 40. Jahrestagung für Akustik (DAGA), 2014.

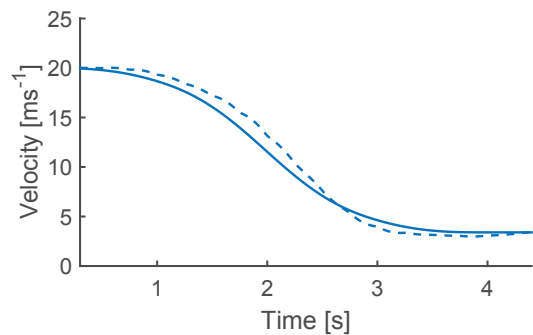


(a) Relativgeschwindigkeit  $\|\vec{v}_{SR}\|$

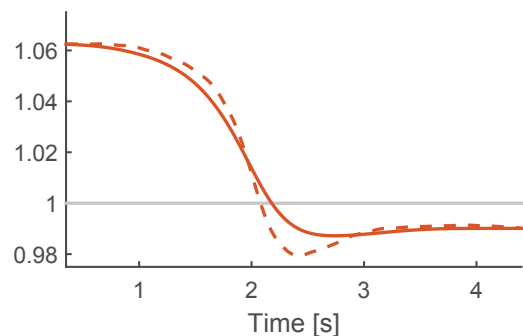


(b) Doppler-Frequenzverhältnis  $f_R/f_S$

**Abbildung 4:** Physikalische Größen für die Catmull-Rom Interpolation der Trajektorie in Abbildung 2.



(a) Relativgeschwindigkeiten  $\|\vec{v}_{SR}\|$



(b) Doppler-Frequenzverhältnis  $f_R/f_S$

**Abbildung 5:** Physikalische Größen für die Prädiktion der DES-Trajektorie ( $\alpha = 0.9$ ) in Abbildung 2 (vorgegebener Verlauf: durchgezogene Linie, Prädiktion: gestrichelte Linie).