

# Bewegungserkennung mit Wearables für Embodied Trainings in Serious Games

Heiko Holz<sup>1</sup>, Benedikt Beuttler<sup>2</sup>, Alexandra Kirsch<sup>3</sup>

Graduiertenschule und Forschungsnetzwerk LEAD, Universität Tübingen<sup>1</sup>

Lehrstuhl für Forschungsmethoden und Mathematische Psychologie, Universität Tübingen<sup>2</sup>

Lehrstuhl für Mensch-Computer-Interaktion, Universität Tübingen<sup>3</sup>

## Zusammenfassung

Embodied Trainings – Übungen mit Ganzkörperbewegungen und Gesten – können das Lernen mit computergestützten Lernspielen für Kinder mit Lernschwächen effizienter und motivierender gestalten. In dem Lernspiel „Silbenschwüngen mit Betonung“ nutzen wir die in Smartphones und Fitnessarmbänder eingebauten Accelerometer zur Bewegungs- und Gestenerkennung. Damit untersuchen wir die Möglichkeit, inwiefern solche Embodied Trainings auch für mobile Serious Games entwickelt werden können. Wir erreichen bei der Klassifizierung von Silbenschwüngen unter Einsatz von Dynamic Time Warping und nur einem Training pro Geste bei Erwachsenen eine Genauigkeit von 99.3 % und bei Kindern zwischen 82.2 % und 93.9 %. Die Ergebnisse zeigen, dass einfache Gesten wie Silbenschwünge zuverlässig erkannt werden können, verdeutlichen jedoch auch die Herausforderung bei der Entwicklung von Embodied Trainings für Kinder.

## 1 Einleitung

In der Lerntherapie können körperbetonte Übungen – sogenannte Embodied Trainings – lernschwache Kinder durch den Einsatz von (Ganzkörper-) Bewegungen zusätzlich fördern. Embodied Trainings zielen auf die Entwicklung körpereigener Gedächtnisstützen ab, bei denen der Lernprozess positiv durch Körperbewegungen und -erfahrungen beeinflusst wird. So kann der gelernte Inhalt später besser abgerufen werden (Dackermann et al., 2017).

Für Kinder mit einer Lese-Rechtschreib-Schwäche (LRS) werden Embodied Trainings in vielen analogen Förderkonzepten erfolgreich eingesetzt. Das Silbenschwüngen ist eine dieser Fördermaßnahmen (Michel, 2008; Reuter-Liehr, 1993). Indem die Kinder den Sprachrhythmus von Worten schwingen (Abb. 1), wird durch ein gezieltes Training der Silbenanalyse eine Verbesserung der Lese- und Rechtschreibleistung erwartet.

Um Embodied Trainings auch in mobilen Serious Games zu unterstützen, können Accelerometer verwendet werden, die in mobilen Geräten verbaut sind. Mit den Accelerometern können körperbasierte Interaktionen wie Bewegung oder Gestik erkannt und klassifiziert werden.

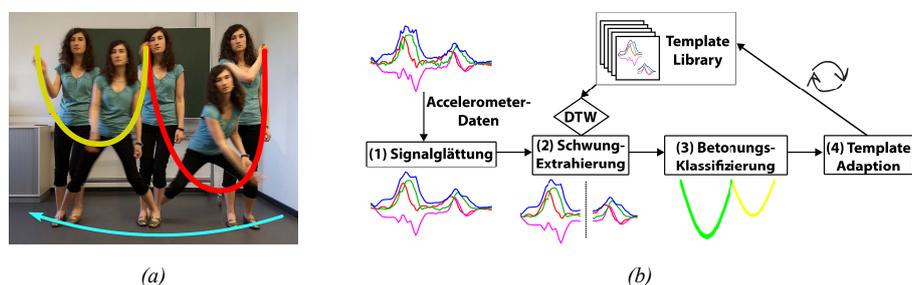


Abbildung 1: Das „Silbenschwingen mit Betonung“ für das Wort „malen“. Kinder lernen durch „Tanzen“ des Sprachrhythmus‘ (Abb. 1a), Wörter zu silbieren und das Betonungsmuster zu bestimmen. Dabei sprechen sie jede Silbe deutlich aus, schreiten gleichzeitig seitwärts durch den Raum und schwingen mit der Schreibhand – für betonte Silben einen größeren, stärkeren und für unbetonte einen kleineren – Silbenbogen von der linken zur rechten Schulter. Abb. 1b zeigt das System zur Klassifizierung der Schwungbewegungen.

Obwohl nahezu perfekten Ergebnisse von bis zu 99.79 % der Gestenerkennung mit solchen Geräten (Xie und Cao, 2016; Akl und Valaee, 2010) dafür sprechen, sind uns keine mobilen Serious Games bekannt, welche durch Gestenerkennung Embodied Trainings unterstützen.

Als Teil des mobilen Serious Games „Prosodiya“ (Holz et al., im Druck) haben wir das Embodied Training „Silbenschwingen mit Betonung“ (SmB) entwickelt (Abb. 1a). Prosodiya fokussiert im Besonderen die Sensibilisierung auf die sprachrhythmische Struktur der Silbenbetonung. Aufgabe der Kinder beim SmB ist es, für vorgegebene Wörter das Betonungsmuster zu schwingen. Die Schwungbewegungen werden mit Accelerometern aufgenommen und durch Gestenerkennung als Betonungsmuster klassifiziert, sodass die Lösung der Kinder automatisch ausgewertet werden kann. Im Gegensatz zum klassischen Silbenschwingen unterscheiden wir zwischen betonten und unbetonten Silben, da Schwierigkeiten im Erkennen der Betonung mit dem Phänomen der LRS assoziiert zu sein scheinen (Sauter et al., 2012).

In diesem Artikel untersuchen wir anhand des SmB, inwiefern Smartphones und Fitnessarmbänder mit eingebauten Accelerometern für die Entwicklung von Embodied Trainings mit einfachen, körperbasierten Interaktionen für mobile Serious Games verwendet werden können.

## 2 Methodik

Für die Datenaufnahme haben wir Android-Smartphones und herkömmliche Fitnessarmbänder verwendet, deren Accelerometer eine Sampling-Rate von 50 Hz, bzw. 10 Hz, besitzen. Wie wir in Nutzertests festgestellt haben, sind Smartphones aufgrund ihrer Größe für Grundschulkinder zum Schwingen ungeeignet. Deshalb wird bei ihnen nur das Fitnessarmband verwendet. Die Extrahierung und Klassifizierung der Schwungbewegungen erfolgt in vier Schritten (Abb. 1b): In Schritt 1 wird das Beschleunigungssignal mittels eines Tiefpassfilters geglättet.

In Schritt 2 werden Schwünge einzelner Silben extrahiert. Ein Schwellenwertverfahren durchsucht das Beschleunigungssignal nach Maxima (Peaks), welche sich zwischen zwei Minima befinden. Die Minima müssen zwischen 500 ms und 1800 ms auseinander liegen. Von diesen Signalausschnitten wird die Ähnlichkeit zu bereits klassifizierten Schwüngen der Template

Library berechnet. Die Berechnung erfolgt mittels Dynamic Time Warping (DTW). DTW bildet zwei zu vergleichende Sequenzen mittels nichtlinearer zeitlicher Verzerrung aufeinander ab (vgl. u. a. Keogh und Ratanamahatana, 2004). Der Vorteil gegenüber statistischen Methoden liegt darin, dass für personalisierte Gesten bei gleicher oder besserer Erkennungsleistung jeweils nur ein Trainingsdatensatz benötigt wird. Bei ausreichender Ähnlichkeit werden die Signalausschnitte als Silbenschwung klassifiziert, ansonsten verworfen.

Da die Kinder sehr unbeständig schwingen – unbetonte Silben werden oft in einem Wort stärker geschwungen als betonte in anderen und umgekehrt – kann die Betonung nicht mit der berechneten Ähnlichkeit zuverlässig bestimmt werden. Deshalb wird in Schritt 3 die Betonung anhand des Peaks bestimmt: Der Schwung mit dem größten Peak innerhalb eines Wortes wird als betont, die anderen als unbetont klassifiziert. Dies wurde gemeinsam mit Lerntherapeuten bestimmt, da alle im Lernspiel verwendeten Wörter nur eine betonte Silbe besitzen.

In Schritt 4 werden korrekt klassifizierter Schwünge zur Template Library hinzugefügt. Da die Schwünge bei längerer Spieldauer schwächer und ungenauer ausgeführt wurden, muss die Template Library mit der Zeit angepasst werden. Sie enthält nur die Schwünge der letzten fünf Minuten, mindestens jedoch fünf Schwünge.

### 3 Evaluation

Das vorgestellte System wurde anhand zweier Nutzertests evaluiert. Für jedes Wort wurden die Accelerometerdaten ab ca. 500 ms vor dem Schwingen bis ca. 500 ms nach dem Schwingen mit einer Android-App aufgenommen. Mit dem ersten Wort wird die Template Library initiiert. Im ersten Nutzertest haben fünf Erwachsene im Alter zwischen 20 und 32 Jahren ( $M = 26$ ,  $SD = 2.7$ ) mit einem Smartphone das SmB durchgeführt. Insgesamt wurden 11 Wörter mit den vier häufigsten Betonungsmustern der deutschen Sprache aufgenommen. Das System erzielte bei der Klassifikation eine Genauigkeit von 99.3%.

Im zweiten Nutzertest wurde das SmB von 9 Kindern im Alter zwischen 10 und 13 Jahren ( $M = 11.1$ ,  $SD = 1.05$ ) mit einem Fitnessarmband durchgeführt. Je nach Motivation wurden zwischen 11 und 14 Wörter aufgenommen. Zu Beginn wurde den Kindern das SmB samt Prinzip der betonten Silbe ausführlich erklärt. Um die korrekte Bestimmung der Betonung sicherzustellen, wurde das SmB zusätzlich vor jedem Wort vorgeführt.

Dennoch konnten wir beobachten, dass einigen Kinder Schwierigkeiten haben, Silben gemäß des Förderkonzepts zu schwingen: Betonte Silben wurden oft schwächer geschwungen als unbetonte oder waren kaum von diesen zu unterscheiden. Es ist nicht klar, ob dies an mangelhaftem Verständnis der Wortbetonung oder an fehlendem Feedback zur Schwungintensität liegt. Deshalb wurden sowohl alle 115 geschwungenen Wörter ( $D1$ , 307 Silbenschwünge) als auch ein bereinigter Datensatz ( $D2$ ) evaluiert.  $D2$  enthält nur die 95 Wörter (258 Silbenschwünge), bei welchen die Kinder das Betonungsmuster korrekt benannt haben und die betonte Silbe stärker oder nur geringfügig schwächer geschwungen wurde ( $Peak_{betont} \geq \max_{Acc} - \Theta_{Peak}$ ).

Ohne Berücksichtigung der Betonung wurden die Silbenschwünge mit einer Genauigkeit von 96.8% für  $D1$  und 97.7% für  $D2$  erkannt. Die Extrahierung der Silbenschwünge mit anschließender Klassifizierung der Betonung erreicht eine Genauigkeit von 83% ( $M = 82.8\%$ ,  $SD = 13.8\%$ -Pkt.) für  $D1$  und 94.6% ( $M = 93.9\%$ ,  $SD = 6.5\%$ -Pkt.) für  $D2$ .

## 4 Diskussion und Ausblick

Das vorgestellte System ist in der Lage, einfache Gesten wie Silbenschwünge zuverlässig zu erkennen. Bei Erwachsenen funktioniert dies mit Smartphones und Fitnessarmbändern. Kinder konnten sich aufgrund der Größe von Smartphones nur mit Fitnessarmbänder uneingeschränkt bewegen. Die Klassifikation der Silbenbetonung ist bei Erwachsenen ausreichend, während sie bei Kindern stark variiert. Die abweichenden Ergebnisse entstehen durch die im Nutzertest festgestellte Schwierigkeit, die Silben gemäß der Anleitung zu schwingen. Dies verdeutlicht drei wichtige Punkte, die bei der Entwicklung sensorgestützter Embodied Trainings für Kinder zu beachten sind: Erstens sollten Fitnessarmbänder aufgrund ihrer Größe und Tragekomfort verwendet werden. Zweitens müssen Systeme an die Bedürfnisse und das Verhalten der Kinder angepasst werden. Drittens darf von den Kindern das korrekte Ausführen der Gesten nicht vorausgesetzt werden, sondern muss verstanden und ausführlich trainiert werden.

Insgesamt konnten wir feststellen, dass Embodied Trainings mit einfachen, körperbasierten Interaktionen für mobile Serious Games entwickelt werden können. Abhängig der Zielgruppe eignen sich unterschiedliche mobile Geräte, in denen ein Accelerometer verbaut ist.

In Zukunft wollen wir die aufgedeckten Limitierungen durch eine bessere und intensivere Trainingsphase überwinden, welche direkt in das mobile Lernspiel integriert wird. Eine Figur aus der Spielwelt soll den Kindern ausführlich das Silbenschwingen erklären und direktes Feedback geben, ob die betonte Silbe ausreichend unterschiedlich geschwungen wurde.

## Literaturverzeichnis

- Akl, A. & Valae, S. (2010). Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, & compressive sensing. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (S. 2270–2273).
- Dackermann, T., Fischer, U., Nuerk, H.-C., Cress, U. & Moeller, K. (2017). Applying embodied cognition: from useful interventions and their theoretical underpinnings to practical applications. *ZDM*, 1–13.
- Holz, H., Beuttler, B., Brandelik, K. & Brandelik, J. (im Druck). Prosodiya - ein Lernspiel zur Förderung des Sprachrhythmus bei Kindern mit LRS. In *Mensch und Computer 2017 - Tagungsband*. Gesellschaft für Informatik.
- Keogh, E. & Ratanamahatana, C. A. (2004). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3), 358–386.
- Michel, H.-J. (Hrsg.). (2008). *FRESCH. Freiburger Rechtschreibschule. Fit trotz LRS*. Buxtehude: AOL-Verlag.
- Reuter-Liehr, C. (1993). Behandlung der Lese-Rechtschreibschwäche nach der Grundschulzeit: Anwendung und Überprüfung eines Konzeptes. *Zeitschrift für Kinder- und Jugendpsychiatrie*, 21(3), 135–147.
- Sauter, K., Heller, J. & Landerl, K. (2012). Sprachrhythmus und Schriftspracherwerb. *Lernen und Lernstörungen*, 1(4), 225–239.
- Xie, R. & Cao, J. (2016). Accelerometer-Based Hand Gesture Recognition by Neural Network and Similarity Matching. *IEEE Sensors Journal*, 16(11), 4537–4545.