
# KI-basiertes Feedback im DaF/DaZ-Ausspracheunterricht? Didaktische Überlegungen zum Einsatz kostenfreier MAPT-Apps

Katrin Engelmayr-Hofmann & Sandra Reitbrecht, Wien

ISSN 1470 – 9570

# KI-basiertes Feedback im DaF/DaZ-Ausspracheunterricht? Didaktische Überlegungen zum Einsatz kostenfreier MAPT-Apps

Katrin Engelmayr-Hofmann & Sandra Reitbrecht, Wien

Betrachtet man didaktische Konzepte für den Ausspracheunterricht, so spielt Feedback eine zentrale Rolle. Der Beitrag nimmt Entwicklungslinien in den Blick, skizziert den Forschungsstand zu KI-Tools im Aussprachetraining und untersucht, wie Feedback von KI-gestützten Systemen (insb. *Mobile Assisted Pronunciation* (MAPT)-Apps) gestaltet wird. Mit einem Analyseinstrument werden exemplarisch zwei kostenfreie KI-basierte MAPT-Tools für Deutsch als Fremd-/Zweitsprache untersucht und Konsequenzen für ihren Einsatz durch Lernende und Lehrende abgeleitet.

## 1. Hinführung – Aussprache im Fremdsprachenunterricht (FSU)

Feedback zu Artikulationsversuchen zielsprachlicher phonetischer Phänomene bildet ein zentrales Element in didaktischen Konzepten für den Ausspracheunterricht (vgl. Hirschfeld & Reinke 2016: 161; Pennington & Rogerson-Revell 2019: 204–205). In ihrem Vorschlag einer Basissequenz für die didaktische Arbeit an einem bestimmten phonetischen Lernziel verorten Hirschfeld & Reinke (2016: 161) die Phase des expliziten Feedbacks *nach* der Hörkontrolle und ersten Imitationsversuchen sowie *vor* dem intensiven Üben des Zielphänomens durch Automatisierung und Anwendung. Das Feedback soll an dieser Stelle des Lern- bzw. Vermittlungsprozesses bei der Artikulation des phonetischen Zielphänomens unterstützen. Dabei hat – in unserem Verständnis – primär eine Orientierung an der Verständlichkeit bzw. an den individuellen Lernzielen der Lernenden zu erfolgen. Ebenso folgen wir Jenkins‘ ressourcen- statt defizitorientiertem Ansatz (2000: 208–212), der das Ziel von (Feedback im) Ausspracheunterricht nicht in einer „accent reduction“, sondern in einer „accent addition“ sieht. Diese soll es Lernenden ermöglichen, ihr Aussprachерepertoire um Aussprachevarianten zu erweitern, ohne dabei andere (lerner:innensprachliche) Aussprachevarianten aufgeben zu müssen.

Belegt ist die Sinnhaftigkeit expliziten Feedbacks im Ausspracheunterricht unter anderem durch Studien, die zeigen, dass Fremdsprachenlernende Schwierigkeiten haben, lerner:innensprachliche, aber auch zielsprachenkonforme Lautproduktionen in der eigenen Aussprache als solche zu erkennen (Dlaska & Krekeler 2008). Hinsichtlich der Frage, wer Feedbackgeber:in sein soll, geht Kelz in seinen zehn Thesen für den

Ausspracheunterricht so weit zu behaupten, dass der „personale Lehrer […] in der Ausspracheschulung durch kein Medium zu ersetzen“ (1992: 24) ist. Evaluierung und Diagnose fallen in seinem Verständnis in das Aufgabengebiet der Lehrenden (vgl. Kelz 1992: 35–37). Ausgeklammert wird durch diese alleinige Relevantsetzung der Lehrperson die Frage, ob es auch Peers gelingen kann, erfolgreich Feedback zu geben. Ebenfalls nicht absehbar waren 1992 die aktuellen Entwicklungen in den Bereichen automatischer Spracherkennung und Künstlicher Intelligenz. Diese haben dazu geführt, dass mittlerweile digitale Lernmedien für das Üben der Aussprache in der Fremd- bzw. Zweitsprache Deutsch mit KI-gestützten Feedbacksystemen vorliegen. Dieser Beitrag nimmt diese Entwicklungslinie in den Blick, zeigt zunächst grundlegende Erkenntnisse zur Relevanz und zu den Gestaltungsmöglichkeiten von Feedback für den Ausspracheunterricht auf und gibt anschließend einen Überblick über den Forschungsstand zu KI-Tools im Kontext von Aussprachetraining. Davon ausgehend greift er die Frage auf, wie Feedback zur Aussprache im Bereich DaF/DaZ von kostenfrei nutzbaren KI-Tools gestaltet wird. Zur Beantwortung dieser Frage wird ein Analyseinstrument zum Feedbackverhalten entwickelt und für die Analyse von zwei gezielt für den DaF/DaZ-Ausspracheunterricht konzipierten kostenfreien KI-gestützten digitalen Lernangeboten eingesetzt, sodass abschließend Konsequenzen für Lernende wie auch Lehrende im Umgang mit diesen Lernangeboten abgeleitet werden können. Der Beitrag schließt mit einem kurzen Ausblick auf den Einsatz KI-basierter Aussprache-Apps im Kontext DaF/DaZ.

## 2. Feedback im Ausspracheunterricht – Wirksamkeitsbelege und Qualitätsmerkmale

Hinsichtlich der Frage nach der Wirksamkeit von Feedback liefert die internationale aussprachedidaktische Forschungsliteratur robuste Evidenz. So können Lee et al. (2015) in einer Meta-Analyse von 86 aussprachedidaktischen Interventionsstudien aufzeigen, dass Interventionen mit Feedback zu größeren positiven Effekten führen als solche ohne. Saiton & Lyster (2012) belegen diesen Vorteil von Feedback in einer direkten Vergleichsstudie von Interventionen mit bzw. ohne Feedback in Form einfacher Recasts, also zielsprachlicher Wiederholungen von lerner:innensprachlichen Aussprachevarianten durch die Lehrperson, ebenfalls.

Dass darüber hinaus auch der Qualität des Feedbacks Bedeutsamkeit für den Lernerfolg zukommt, veranschaulichen Dlaska & Krekeler (2013) in einem direkten Vergleich von *implizitem* und *explizitem* Feedback. Dabei legen sie ein dreischrittiges Verständnis von Feedback zugrunde, welches Wissen (1) über den eigenen Lernstand und (2) über die Lernziele sowie (3) Angebote und Informationen beinhaltet, wie man diese Lernziele erreichen kann. *Implizites* Feedback (ähnlich dem Recast bei Saito & Lyster 2012) erhielten die Lernenden in dieser Studie, indem sie sowohl ihre eigene Aussprache als auch eine Modellaussprache vergleichend hören konnten. *Explizites* Feedback war qualitativ gestaltet und beinhaltete zudem auch die dritte Komponente, umfasste also verbale Erklärungen zu den individuellen Ausspracheabweichungen und zu den Zielregeln, ergänzt um Beispiele (vgl. Dlaska & Krekeler 2013: 29), die einen Weg zur Zielerreichung aufzeigen sollten. Die Ergebnisse der Studie zeigen, dass auch ein solches individualisiertes Feedback nicht zwingend bei allen Lernenden zu einer Verbesserung der Aussprachеleistung führt. Dennoch ist der Anteil an Personen, die davon profitieren, deutlich höher als der Anteil an positiven Ausspracheentwicklungen in der Gruppe, die ausschließlich implizites Feedback erhielt.

In ähnlicher Weise können die Ergebnisse von Huang et al. (2023) interpretiert werden. Sie zeigen, dass vollständiges Lehrendenfeedback mit Hinweisen zur Zielerreichung nicht nur unmittelbar, sondern auch längerfristig zu besseren Verständlichkeitsbewertungen führte. Peer-Feedback hingegen, welches sich in seiner Qualität unterschiedlich gestaltete und oftmals keine Hinweise zur Zielerreichung lieferte (vgl. Huang et al. 2023: 57–58), führte nur unmittelbar nach der Peer-Lernsituation zu einer besser verständlichen Aussprache, nicht aber längerfristig. Damit unterstreichen auch diese Ergebnisse die Relevanz der Qualität von Feedback für den Lernerfolg.[1]

Effektives Feedback im Ausspracheunterricht geht also über eine einfache Form von Recasts hinaus und ist „mit entsprechenden Methoden und Hilfsmitteln“ (Hirschfeld & Reinke 2016: 161; siehe auch Dieling & Hirschfeld 2000) didaktisch angereichert. Diese dienen nicht nur der Bewusstmachung der lerner:innensprachlichen sowie der zielsprachlichen Artikulationsvarianten, sondern unterstützen Lernende auf dem Weg der „accent addition“. Neben den bei Dlaska & Krekeler (2013) gezielt überprüften

[1] Diese Ergebnisse sollen zugleich aber nicht als eine Absage an Peer-Feedback im Ausspracheunterricht gelesen werden. Besonders fruchtbar kann nämlich das Feedbackgeben für Lernende sein (s. Martin & Sippel 2023), welches in der Studie von Huang et al. (2023) nicht näher untersucht wurde.

verbalen Erklärungen schlagen Hirschfeld und Reinke (2016: 158) in diesem Zusammenhang unter anderem auch Visualisierungen, Verfahren der Ableitung des Ziellautes von anderen (bereits bekannten) Lauten oder Geräuschen, den Einsatz unterstützender Gesten und Mitbewegungen (z. B. Hüpfen, Klatschen) oder Summen, Brummen und lautmalerisches Singen vor (s. u.a. Dieling & Hirschfeld 2000).

Für Visualisierungen als Elemente von Feedback zeigen Bliss et al. (2018) die Vielfalt technischer Möglichkeiten auf und gruppieren diese in drei Formen computergestützten visuellen Feedbacks auf lerner:innensprachliche Artikulationen: (1) Feedback in Form von Visualisierungen akustischer Parameter (z. B. Spektrogramme), die keine Darstellungen der artikulatorischen Positionen oder Bewegungsabläufe beinhalten; (2) Feedback in Form von bildgebenden Verfahren (z. B. Ultrasound), die akustische Informationen für Visualisierungen artikulatorischer Positionen und Bewerbungsabläufe heranziehen; (3) Feedbackformen, die nicht die Artikulation der Lernenden zeigen, aber VR-/AR-gestützte Visualisierungen[2] von Zielartikulationen anbieten. Der von Bliss et al. (2018) aufgearbeitete Forschungsüberblick verdeutlicht allerdings, dass bisher divergierende Befunde zur Wirksamkeit der beiden erst genannten Feedbackverfahren vorliegen. Sehr wohl aber bieten die analysierten Studien Indizien für die Relevanz folgender Qualitätsmerkmale von Feedback über die konkrete Form der Visualisierung hinaus: seinen logischen Charakter, Verständlichkeit, Unmittelbarkeit, Förderung von Artikulationsvergleichen und Flexibilität. Zudem sollte technologiegestütztes Feedback einen Mehrwert gegenüber anderen Feedbackformen haben und für Lehr-Lern-Kontexte „leistbar“ sein (vgl. Bliss et al. 2018: 11–14).

Bliss et al. (2018) unterstreichen damit, dass vermutlich nicht die konkrete technologische Entwicklung zu einer Verbesserung des Ausspracheunterrichts beitragen kann, sondern dass diese immer im Kontext des gesamten didaktischen Designs zu sehen ist und hier auch weitere Qualitätskriterien greifen. Diese Schlussfolgerung findet eine Bestätigung in der oben zitierten Meta-Studie von Lee et al. (2015). Durch entsprechende Moderatoranalysen ließ sich herausarbeiten, dass sich das von Lehrenden gegebene Feedback als effektiver als die technologie- bzw. computergestützte Rückmeldung erwies. Eine mögliche Erklärung dieses Unterschieds sehen Lee et al. (2015) darin, dass entsprechende Tools und Computerprogramme weniger genaue Einschätzungen der Aussprechperformanz vornehmen, einen geringen Grad an Adaptivität erreichen

und zudem weniger gut in der Lage sind, angemessenes Feedback zu geben (vgl. Lee et al. 2015: 361), wobei im „Angemessenheitsbegriff" die in diesem Abschnitt herausgearbeiteten Qualitätsmerkmale von Feedback angesetzt werden können. Hieraus ergibt sich die Frage, inwieweit digitale Anwendungen oder im weiteren Sinne Tools basierend auf sogenannter Künstlicher Intelligenz, die als eine nächste Stufe technologischer Entwicklung verstanden werden können, als „Feedbackinstanz" für den Ausspracheunterricht agieren (können).

## 3. Appbasiertes Aussprachetraining

Betrachtet man die technologischen Entwicklungen der vergangenen Jahrzehnte, deren Eingang in unterrichtliche Kontexte nicht unwesentlich durch die Covid-19-Pandemie geprägt wurde, so lässt sich ein steter Anstieg KI-basierter Tools ausmachen, die Nutzende beim Sprachhandeln und -lernen auch im Bereich der Aussprache unterstützen können.[3] Unterscheiden lassen sich dabei zwei Arten: (1) Applikationen, die nicht explizit für ein Aussprachetraining entwickelt wurden, deren Funktionen jedoch ein Aussprachetraining grundsätzlich zulassen, und (2) Applikationen, die ein Aussprachetraining explizit zum Ziel haben.

### 3.1 Aussprachetraining mit nicht dafür konzipierten Apps

Zu den ersteren Anwendungen (1) können omnipräsente maschinelle Übersetzungstools (im Weiteren MT-Tools) – wie *Google Translate* oder *DeepL* –, aber auch Online-Wörterbücher sowie Textverarbeitungsprogramme gezählt werden. Einzelne, teilweise KI-basierte, Funktionen dieser Tools ermöglichen es Lernenden einer Fremdsprache sich auf einfache Weise eine (normierte und standardsprachlich-orientierte) Aussprache individueller Wörter bis hin zu ganzen Sätzen ausgeben zu lassen (*Text-to-Speech;* kurz: TTS). Die meist mittels Lautsprechersymbol abspielbaren Hörbeispiele orientieren sich an einer Standardvarietät, wobei Variation auf Ebene der Standardsprache kaum abgebildet wird. Dies kann auch für das Deutsche festgehalten werden – weder bei *Google Translate* noch bei *DeepL* lassen sich Standardvarietäten des Deutschen (bislang, Stand

---

[2] VR steht für *virtual reality*, AR für *augmented reality*.

[3] Da der vorliegende Beitrag explizit Aussprache fokussiert, werden Untersuchungen zur Verwendung KI-basierter Tools innerhalb mündlicher fremdsprachlicher/mehrsprachiger Kommunikationssituationen ausgeklammert. Ausgewählte Literatur zu diesem Thema: Shadiev et al. (2019), Ross et al. (2021).

November 2024) finden. Ausnahmen bilden hingegen das amerikanische und das britische Englisch (z.B. bei *Google Translate*) sowie Portugiesisch und brasilianisches Portugiesisch (z. B. bei *DeepL*). *Google Translate* bietet zudem bei Übersetzungen aus oder in beide(n) englischsprachige(n) Varietäten mit Klick auf das Lautsprechersymbol direkt die Möglichkeit, deren Aussprache gezielt zu üben. Weitergeleitet zur *Google*-Suche findet man dort die Möglichkeit eines wiederholten Abspielens des Wortes, das auch „langsamer" erfolgen kann; eine lautsprachliche Vergleichsbeschreibung („Klingt wie"), die das auszusprechende Wort visuell in Silben teilt und eine Betonung mittels Fettsetzung markiert; sowie ein Video, bei dem ein animierter Mund die Lippenbewegungen nachvollzieht. Bei anderen Sprachen[4] existiert bislang keine Verknüpfung zwischen den Tools.

Gerade wohl aufgrund ihrer Einfachheit in Zugang und Anwendung wird der Aussprachefunktion bei MT-Tools in Befragungen lernendenseitig ein großer Nutzen attestiert (vgl. Ata & Debreli 2021: 110; Niño 2020: 21) und Lernende geben an, mit Hilfe von MT-Tools selbstgesteuert die eigene Aussprache zu überprüfen (u.a. Jin & Deifell 2013; Organ 2023; Rangsarittikun 2022; Wirantaka & Fijanah 2021). Inwiefern sie die Tools abseits didaktischer Einbindungen nutzen, um gezielt die eigene Aussprache zu trainieren und auf Dauer zu verbessern (Sprachlernstrategie) oder darauf zurückgreifen, um in einer akuten Situation Kommunikation und Teilhabe aufrecht zu erhalten bzw. zu generieren (Sprachgebrauchsstrategie), bleibt jedoch unklar (vgl. Rangsarittikun 2022: 12). Ergebnisse von Untersuchungen, die eine Vermittlung bestimmter phonetischer Phänomene auf Basis von TTS zum Fokus haben, lassen jedenfalls positive Implikationen (v.a. aufgrund des erhöhten Inputs in der Zielsprache, einer intensiveren Auseinandersetzung mit Graphem-Phonem-Beziehungen und autonomen und räumlich-zeitlich flexiblen Lernens) für ein Aussprachetraining zu (u. a. Cardoso 2018: 21; He & Cardoso 2021).

Zusätzlich zu den genannten Funktionen können Textverarbeitungsprogramme, Diktierprogramme und MT-Tools mittels einer Verschriftlichung mündlicher Eingaben (*Speech-to-Text,* kurz: STT – oft *Automatic Speech Recognition*, kurz: ASR) eine indirekte Überprüfung der eigenen Aussprache ermöglichen. Sie können dabei zu

[4] Für Deutsch lassen sich ein verlangsambares Aussprachebeispiel und eine Vergleichsbeschreibung nur mittels gezielter Google-Suche (Prompt: „pronunciation WORD german") abrufen – jedoch ohne Video, das animierte Lippenbewegungen zeigt.

positiven Emotionen ob einer (An-)Erkennung der eigenen (korrekten) Aussprache durch die Technologie (s. Kennedy 2021) sowie, sofern didaktisch eingebunden, zu einer nachweislichen Verbesserung der Aussprache auf segmentaler Ebene (u.a. McCrocklin 2019a, 2019b) führen.

Studien, die einen Einsatz von MT zu Aussprachetrainingszwecken untersuchten (u. a. van Lieshout & Cardoso 2022; Niño 2025; Papin & Cardoso 2022, 2023, 2025), legen ebenfalls nahe, dass die kombinierte Nutzung der TTS- und STT-Funktionen in *Google Translate* Lernenden ein selbstgesteuertes Aussprachetraining ermöglichen kann – vorausgesetzt, diese verfügen über ausreichend phonologisches Wissen und nutzen *Google Translate* phänomengeleitet (vgl. Papin & Cardoso 2022: 326).

Andere Studien zeichnen allerdings ein kritischeres Bild hinsichtlich der ASR-Akuratheit von STT-Funktionen und ihrer Eignung für Aussprachelernzwecke. Sie berichten davon, dass die Aussprache vor allem von Lernenden, die gerade anfangen eine Sprache zu lernen, Gefahr laufe, von der Technologie nicht erkannt zu werden (vgl. Ashwell & Elam 2017: 71–72; Medvedev 2016: 191; Niño 2020: 11). Dieses Ergebnis hat sich vermutlich inzwischen durch die flächendeckende Verwendung neuronaler Netzwerke insofern verändert, als dass nun der umgebende Kontext der KI „verrät“, welche Wörter und Sätze ausgesprochen wurden und der Algorithmus entsprechende Korrekturen automatisiert durchführt. Es gilt also kritisch zu hinterfragen, inwiefern Technik überhaupt darüber entscheiden sollte, wie und ob Lernendensprache verstanden wird und welche Aussprache(-variante) bzw. welches Ausspracheziel Lernende verfolgen sollten.

Obwohl nicht für ein gezielte Aussprachetraining konzipiert, bieten zahlreiche gängige Tools demnach aus Sicht der Nutzenden und einiger Forschenden auf Basis ihrer Funktionalität Möglichkeiten autonomer Selbstaneignung und -überprüfung von Aussprache. (Unendliche) Wiederholbarkeit sowie die zeitliche wie personelle Entkopplung von der Lehrperson scheinen dabei entscheidende Faktoren für Lernende zu sein, diese selbstgesteuert für das eigene Aussprachetraining einzusetzen (vgl. Bin Dahmash 2020: 234). Die hier offenbarten Nutzungsgründe entsprechen damit auch jenen vier Vorteilen, die Neri et al. bereits 2002 in Bezug auf computergestützte Aussprachetrainings (engl. *computer-assisted pronunciation training,* CAPT) formulierten: 1. Individuelle Anpassungsmöglichkeiten an die Bedürfnisse der Lernenden 2. grenzenloses und selbstbestimmtes Lernen, 3. Reduktion von (Sprech-)Ängsten sowie 4. Möglich-

keiten eines sofortigen, direkten Feedbacks zur Ausspracheperformanz (vgl. Neri et al. 2002: 450). Zugleich zeigen aber die mit der Tool-Nutzung einhergehenden Probleme einer fehlenden Repräsentation sprachlicher Variation, eines fehlerbehafteten Erkennens von Lerner:innenaussprache (abseits der KI-inhärenten Kontext-/Syntaxanalyse) sowie einer nutzer:innenseitigen Notwendigkeit um phonologisches Wissen auf, dass das technologiegestützte Feedback ohne didaktische Einbettung in dieser Form bisher kaum einen Mehrwert gegenüber herkömmlichen Methoden zu leisten vermag (s. Bliss et al. 2018; Lee et al. 2015).

### 3.2. MAPT-Apps

Im Vergleich zu den unter (1) genannten Tools haben (2) MAPT-Apps (engl. *mobile-assisted pronunciation training*-Apps) ein Aussprachetraining zum expliziten Ziel. Sie enthalten übergreifend oft eingebettete Elemente wie Audiodateien; eine Möglichkeit zur Audioaufnahme und -wiedergabe über ein Mikrofon; eine automatische Sprach-erkennung (ASR) inklusive unterschiedlichen Feedbackformen; Videodateien bzw. Verweise auf externe Videos; eine Möglichkeit zur Videoaufnahme per Kamera sowie die physische Interaktion mit der App, zum Beispiel über Touchscreen, Tastatur [QWERTZ/Y und IPA] oder Beschleunigungsmesser (vgl. Kaiser 2018: 2–3). MAPT-Apps bieten damit grundlegend verschiedene Optionen, Aussprache gezielt(er) einzu-üben, und können – gerade auch im Vergleich mit CAPT – positive Auswirkungen auf die Motivation haben (vgl. Lan 2022). Verschiedene Studien konnten für den Bereich EFL zudem bereits nachweisen, dass ein gezielter Einsatz von MAPT im fremd-sprachlichen Unterrichtskontext eine (wenngleich nicht immer signifikante) Steigerung beim phonologischen bzw. phonetischen Hören wie auch in der Aussprache der Zielsprache bewirkt: Dies zeigten zum Beispiel Untersuchungen auf segmentaler Ebene zu /æ ɑː ʌ ə/ und der Unterscheidung /s – z/ unter 52 spanischen EFL-Studierenden (vgl. Fouz-González 2020) sowie zur Rezeption und Produktion von /o/ und /ɔ/ bei 24 koreanischen EFL-Lernenden (vgl. Lee 2021).

MAPT-Apps verfolgen dabei in den meisten Fällen ein sequenzielles Vorgehen. Mit Blick auf den segmentalen Bereich setzen sie zum Beispiel der eigentlichen Aus-spracheproduktion Lautdiskriminierungsübungen voran (s. Hirschi et al. 2022). Erst danach wird die eigene Aussprache aufgezeichnet und basierend auf ASR bewertet. Das Feedback besteht zumeist aus einer kategorialen Bewertung (richtig, falsch). Bei

fehlerhafter Aussprache werden weitere Audiobeispiele inklusive visueller Einfärbungen des Ziel-Phonems auf der Textebene für weitere Übungsaktivitäten bereitgestellt (vgl. Hirschi et al. 2022: 4).

Auf suprasegmentaler Ebene können MAPT-Apps wie *Study Intonation* (CDK) zum Beispiel den Tonhöhenverlauf der Nutzenden visualisieren und im Feedback kontrastiv einer Modellvisualisierung gegenüberstellen (vgl. Tan 2021: 4). Statt einer rein kategorialen Bewertung erhalten Lernende hier zudem ein kontrastives und graduelles Feedback, das jedoch von ihnen selbst dahingehend interpretiert werden muss, welche Schritte zu einer Verbesserung beitragen können (vgl. Tan 2021: 6).

Einen Überblick über verschiedene MAPT-Apps, die eine Ausspracheschulung zum Ziel haben, bieten Studien wie jene von Kaiser (2018) oder Walesiak (2017). Ihre Untersuchungsergebnisse, basierend auf Analysen verschiedener iOS- beziehungsweise Android-basierter Aussprache-Apps[5] im Kontext EFL, lassen sich mit Meisarah wie folgt zusammenfassen: Die Mehrzahl der MAPT-Apps konzentrieren sich auf das Vermitteln und Einüben einer Reihe bestimmter segmentaler Merkmale, zum Beispiel mittels IPA-Transkription, einzelner Wörter, Minimalpaare und detaillierter Anweisungen. Teilweise enthalten die Apps visuelles Feedback in Form von Modellvideos oder lassen eine eigene Videoaufnahme der Lippenbewegungen zu Vergleichszwecken zu. Nur sehr wenige Anwendungen hingegen können als Quelle für das Training suprasegmentaler Merkmale wie Melodieverläufe, Rhythmus oder Akzentuierung herangezogen werden (vgl. Meisarah 2020: 73).

Als Hauptproblem der Anwendungen identifiziert Kaiser in seinen Analysen jedoch die Art des Feedbacks. Folgende Formen von Feedback(elementen) lassen sich unserer Einschätzung nach aus seiner Darstellung ableiten:

- Kein Feedback
- Richtig oder falsch (rot oder grün) ohne Erläuterungen
- Selbstgesteuerter Abgleich (audio):
  Eigene Audioaufnahme mit Playbackfunktion, teilweise jedoch mit Zeitlimit

[5] Kaiser (2018) untersuchte insgesamt 105 kostenpflichtige wie kostenfreie iOS-Aussprache-Apps, von denen er 30 in dem oben genannten Beitrag detailliert bespricht. Walesiak (2018) hingegen analysiert 70 frei zugängliche Android-Ausspracheapps, von denen sie 15 bespricht.

- Selbstgesteuerter Abgleich (audiovisuell):
  Eigene Videoaufnahme (Front-Kamera) und gegebenenfalls paarweiser Vergleich mit Modellvideo der Lippenbewegung
- Feedback zur eigenen Aussprache eines Vokals mittels Lokalisation innerhalb einer visualisierten Vokaltabelle (*formant chart*)
- Audio-Visual-Bio-Feedback:
  Darstellung unter anderem von Verlauf und/oder Tonhöhe des Zielwortes vs. der Lerner:inneneingabe in Diagrammform

(in Anlehnung an Kaiser 2018: 3–5)

Keine der genannten Feedbackformen enthält eine genaue Erläuterung dessen, was verändert werden soll, um ein besseres bzw. gemäß Technologie „richtig(er)es" Ergebnis der eigenen Aussprache zu erzielen – ein Ergebnis, das mit Blick auf die in Abschnitt 2 ausgeführten Untersuchungsergebnisse aus pädagogischer Sicht zumindest Fragen aufwirft. Kaiser selbst schlussfolgert deshalb mit Verweis auf Neri et al. (2003: 1159), dass sinnvolles Feedback zu geben bedeute, Feedback zu geben, das von Lernenden interpretiert werden kann – womit möglicherweise eine weitere Vereinfachung der Aufgabe und der Art und Weise, wie Feedback gegeben wird, einhergehe (vgl. Kaiser 2018: 5). Anhand der im Zusammenhang mit Feedback geäußerten Kritik Kaisers werden darüber hinaus auch grundsätzliche Herausforderungen KI-basierter Anwendungen deutlich – gerade mit Blick auf deren Akkuratheit[6]. Für den DaF/DaZ-Kontext wirft dies an erster Stelle die Frage auf, auf Basis welcher Datenkorpora und somit auch auf Basis welcher zielsprachlichen Varietäten (und von wem) die den KI-Anwendungen zugrundeliegenden Algorithmen trainiert werden (u.a. Moorkens 2022). In von Mehrsprachigkeit geprägten Gesellschaften scheint eine Orientierung und ein Messen zweit- und fremdsprachlicher Aussprachekompetenz an einer fiktiven *Native-Speaker*-Norm, die durch eine KI auf statistische Wahrscheinlichkeiten reduziert wird, jedenfalls wenig zielführend.

[6] Siehe Cuccharini et al. (2007).

# 4. Methodisches

## 4.1 Entwicklung und Aufbau des Analyseinstruments

Anhand der theoretischen Auseinandersetzung mit Feedback im Ausspracheunterricht lassen sich übergreifende Kriterien für die Analyse und Bewertung der Feedbackfunktionen von MAPT-Apps für DaF/DaZ ableiten. Diese haben wir in einem Analysebogen zusammengefasst und heuristisch um weitere Analysekriterien ergänzt (s. Anhang 1). Das Analyseinstrument ermöglicht zunächst eine Dokumentation *allgemeiner Informationen zur Applikation* (0) und ist im weiteren Aufbau, basierend auf den drei Schritten von Dlaska und Krekeler (2013), in die Bereiche *Wissen über den Lernstand* (1), *Wissen über das Ziel* (2) sowie *Wissen über Wege zum Ziel* (3) gegliedert. Geleitet durch eine übergreifende Frage berücksichtigt jeder Bereich verschiedene Feedbackmodi und enthält konkrete Beispiele zur Veranschaulichung. Wo möglich, sind direkte Verweise zu Studien oder Quellen gesetzt, die empirische Evidenz zur Relevanz des entsprechenden Kriteriums liefern bzw. referieren.

**0 Allgemeines**
Name:
Produktionsfirma:
Zugang: ✓ ✗ iOs ✓ ✗ Android ✓ ✗ browserbasiert
Auszüge Selbstbeschreibung: ZITATE Webseiten
✓✗ Offenlegung zugrundeliegender Sprachnormen
✓✗ Offenlegung der Art/Weise der Bewertung von Lernendenspracheingaben
Folgende phonetische Phänomene können geübt werden:
✓✗ Segmentalia (Konsonanten, Vokale, etc.) in jeweils einzelnen Übungen
✓✗ Suprasegmentalia (Wortakzent, Melodiebewegungen, Rhythmus, etc.) in jeweils einzelnen Übungen
✓✗ Kombination mehrerer Phänomene in einer Übung

Abb. 1: Kernbereich 0 Allgemeines zum Analyseinstrument

Für eine einleitende *allgemeine Beschreibung* der MAPT-App (siehe Abb. 1) umfasst das Analyseinstrument neben den Informationen zum Namen der App, zur Produktionsfirma und zum Zugang auch Raum für Beschreibungen des Angebots sowie für die Aussprachedidaktik relevante Aspekte: Angaben zu den normativen Referenzen in der App und zur Transparenz hinsichtlich der Bewertung der Lernendenspracheingaben sowie die Phänomenbereiche der Aussprache des Deutschen, die (einzeln oder kombiniert) geübt werden können.

Der Kernbereich 1 des Analyseinstruments (siehe Abb. 2) nimmt in einem nächsten Schritt unter der Leitfrage *Welche Art(en) von Rückmeldung(en) zum Lernstand erfolgen über die App?* gezielt das über die App und ihr Feedbackverfahren vermittelte Wissen über den eigenen Lernstand in den Blick.

| **Modus?** | **Art?** | **Beispiel?** | **Quellen** |
|---|---|---|---|
| schriftsprachlich | kategorial | *richtig, falsch* | Huang et al. (2023)<br>*Kaiser (2018)* |
| | graduell | *Prozentzahlen, „Nah dran!“, „Fast!“* | |
| | qualitativ | *gegenstandsbezogen: „Dein U klingt wie ein I.“* | Huang et al. (2023) |
| visuell | kategorial/graduell | *Farben, Icons* | |
| | qualitativ | *Hervorhebungen im Text (fett, kursiv, Silben, Rhythmus, Textanimation)* | Hirschi et al. (2022) |
| | | *Spektrogramm* | Bliss et al. (2018)<br>*Kaiser (2018)* |
| | | *Melodiekurve (f0)* | Bliss et al. (2018)<br>Tan (2021) |
| | | *visuelle Verortung in einer Lauttabelle (formant chart)* | Bliss et al. (2018)<br>*Kaiser (2018)* |
| auditiv | kategorial/graduell | *Soundeffekte* | |

Abb. 2: Kernbereich 1 Wissen über den Lernstand

Unterschieden wird dabei zwischen verschiedenen Rückmeldungsmodi (schriftsprachlich, visuell, auditiv), welche weiter hinsichtlich der Art und Qualität der Rückmeldung (kategorial, graduell, qualitativ) ausdifferenziert werden. Schriftsprachliche Rückmeldungen haben dabei nicht nur das Potenzial, die lerner:innensprachliche Lautproduktion kategorial oder graduell an der Norm zu messen, sondern diese auch qualitativ zu beschreiben. Im Bereich visueller Rückmeldungen zum Lernstand werden neben visuellen Mitteln wie Farben oder Icons vor allem auch die in der Fachliteratur besprochenen Visualisierungsmöglichkeiten akustischer Analysetools gelistet.

Für den Kernbereich 2 *Wissen über das Ziel* (siehe Abb. 3) umfasst das Analyseinstrument entlang der Fragestellung *Wie wird das phonetische Zielphänomen in der Ausspracheübung dargestellt?* die im Folgenden dargestellten Kriterien.

| **Modus?** | **Beispiel?** | **… der Spracheingabe?** | **Quellen** |
|---|---|---|---|
| schriftsprachlich | Transkriptionszeichen (IPA-Symbole) | VOR / NACH | Huang/Lee/Ballinger (2023) |
| | Beschreibung des phonetischen Zielphänomens | VOR / NACH | |
| visuell | Hervorhebungen im Text *(fett, kursiv, Silben, Rhythmus, Textanimation)* | VOR / NACH | |
| | Markierungen *(Wellen, Pfeile für Melodiebewegungen etc.)* | VOR / NACH | |
| | Abbildungen (konkret) *(Bild von Lippenrundung etc.)* | VOR / NACH | |
| | Spektrogramm (abstrakt) | VOR / NACH | |
| audiovisuell | Artikulationsmodell (Video) *(Art des Videos, Repräsentation von Diversität (Alter/Geschlecht Sprecher:innen, Registervariationen, phono-stilistische Variationen))* | VOR / NACH | |
| auditiv | rein auditives Aussprachemodell (Audio) *(Repräsentation von Diversität (Alter/Geschlecht Sprecher:innen, Registervariationen, phono-stilistische Variationen))* | VOR / NACH | |

Abb. 3: Kernbereich 2 Wissen über das Ziel

Zentral sind auch hier die Modi der Rückmeldung, wobei diese teilweise noch weiter ausdifferenziert werden. Ebenso zentral erscheint im Umgang mit den MAPT-Apps die Frage, ob diese Darstellungen zum Zielphänomen vor und/oder nach der Spracheingabe erfolgen. Daher ist für diesen Aspekt eine eigene Spalte im Analyseinstrument vorgesehen.

| **Modus?** | **Beispiel?** | **Quellen** |
|---|---|---|
| Implizit | vergleichendes Sehen (konkrete Darstellung) *(eigene Videoaufnahme vs. Modellvideo)* | Saito/Lyster (2012) |
| | vergleichendes Sehen (abstrakte Darstellung) *(eigenes Spektrogramm vs. Zielspektrogramm)* | |
| | vergleichendes Hören *(eigene Audioaufnahme vs. Modellaudio)* | Dlaska/Krekeler (2013) Saito/Lyster (2012) |
| implizit adaptiv | (erneutes) Artikulationsmodell (Video) *(mit Überbetonung, Verlangsamung, unterstützenden Gesten etc.)* | Bliss/Abel/Gick (2018) |
| | (erneutes) rein auditives Aussprachemodell (Audio) *(mit Überbetonung, Verlangsamung etc.)* | Huang/Lee/Ballinger (2023) |
| explizit | Erläuterungen/Anweisungen (schriftlich, visuell, auditiv), um Ziel zu erreichen *(Ableitung von anderen Lauten/Geräuschen, Anleitung zur Artikulation/prosodischen Gestaltung)* | Dlaska/Krekeler (2013) Hirschfeld/Reinke (2016) |
| explizit adaptiv | Möglichkeit der Wiederholung *(direkt (auf Empfehlung) oder indirekt durch Wiederholung der Übung (nachgelagert))* | Hirschi et al. (2022) |

Abb. 4: Kernbereich 3 Wissen über Wege zum Ziel

Der Kernbereich 3 *Wissen über Wege zum Ziel* (siehe Abb. 4) schließlich fokussiert auf den dritten zentralen Schritt erfolgreichen Feedbacks, der auch als Feedforward bezeichnet wird und entlang der Leitfrage *Welche Arten von Unterstützung zur Aussprachеverbesserung werden gegeben?* Lösungs- und Lernwege für die Artikulation des Zielphänomens in den MAPT-Apps ermittelt.

Dabei werden auch implizite Formen der Rückmeldung berücksichtigt, die noch keine konkreten Lösungswege anbieten, aber es den Lernenden zumindest ermöglichen, die eigene sowie die Zielproduktion sehend oder hörend zu vergleichen. Ebenso denkbar ist aber, dass auditive oder (audio-)visuelle Modelle über die Zielsprache nach der Spracheingabe in adaptierter Form (also gezielt angepasst an die Bedürfnisse einzelner Lernender) dargeboten werden. Darüber hinaus sind mit Blick auf die Verständlichkeit des Feedforward und die Lernzielerreichung vor allem aber das Vorhandensein expliziter Rückmeldungen wie Erläuterungen und Anweisungen zum Erreichen der Zielartikulation sowie Möglichkeiten der Übungswiederholung unter Anweisung in den MAPT-Apps in der Analyse der Tools zu ermitteln, worauf die letzten beiden Zeilen im Kernbereich 3 des Analyseinstruments abzielen.

### 4.2 Analysierte MAPT-Apps

Die Auswahl der MAPT-Apps basierte auf vier Kriterien: (1) Die Anwendung nennt Aussprachetraining nach eigenen Angaben als Hauptziel (MAPT-App) und ermöglicht dieses für die Zweit-/Fremdsprache Deutsch, wobei (2) der Zugang weitgehend kostenfrei und damit niederschwellig über eine App oder den Browser ermöglicht wird, (3) eine Audioeingabe per Mikrofon erfolgen kann und (4) Feedback nach Aussagen der Produktionsfirmen KI-basiert erfolgt.[7]

Um geeignete Anwendungen zu identifizieren, wurden im Februar 2024 sowohl der *Google Play Store* und *Apple Store* als auch gängige Online-Suchmaschinen (Suchwort „Aussprache App Deutsch“ bei *Google*, *Bing* und *Yahoo*) durchsucht und abgeglichen. Auf dieser Basis wurden zwei Anwendungen identifiziert, von denen jedoch nur die erste vollständig kostenfrei zur Verfügung steht:

1. der webbasierte *Online Aussprachetrainer* des *Goethe Instituts*, der sowohl Informationen und Tipps zu Ausspracheregeln der deutschen Sprache enthält als auch Möglichkeiten der Übung mit direktem Feedback bietet (Goethe-Institut o. J.b),
2. die „KI-gestützte Aussprache- und Sprachlern-App“ *Sylby*, die sich in ihrem Quelltext selbst als „[t]he leading pronunciation training app developed by linguists and AI experts to help you express yourself confidently in foreign languages” (Sylby o. J.) bezeichnet.

### 4.3 Analysemethodisches Vorgehen

Beide Anwendungen stellen klassische MAPT-Apps dar, die Aussprachetraining in den Fokus stellen. Wir testeten diese zunächst unabhängig voneinander und untersuchten sie anhand der Kernbereiche 0 bis 3 des Analyseinstruments hinsichtlich ihres Feedbacks näher. Die Ergebnisse wurden anschließend abgeglichen sowie eine konsensuale Entscheidung hinsichtlich des Analyseergebnisses getroffen. Zudem wurden beide An-

[7] Während es durchaus Anwendungen gibt, die die von Kaiser (2018) und Walesiak (2017) beschriebenen Feedbackfunktionen (z.B. audiovisuell, Darstellung von Tonhöhe/-verlauf) enthalten, beziehen sich diese, zum Zeitpunkt der Sampleerstellung, nur auf das Üben und Erlernen englischer Aussprache und sind damit für die vorliegende Untersuchung ungeeignet. Ebenfalls ausgeschlossen wurden kostenpflichtige Apps, die zeitversetztes, personalisiertes Feedback von L1-Sprecher:innen anbieten (z.B. *Speechling* oder *Busuu*), sowie solche, die Aussprache als einen Nebenaspekt behandeln (z.B. *Babbel*, *Duolingo* oder *Talkpal AI*).

wendungen im Sommersemester 2024 im Rahmen einer 60-minütigen Masterseminareinheit (Seminar Digitale Kompetenz im DaF/DaZ-Unterricht) an der Universität Wien analysiert und getestet. Die teilnehmenden acht Studierenden – hierfür in vier Gruppen (zwei Gruppen pro Tool) aufgeteilt – ergänzten nach einer kurzen Einführung ihre Analyseergebnisse und Eindrücke mittels Kommentarfunktion digital im Kriterienraster. Die Daten aus diesen weiteren Testungen dienten der Perspektivenerweiterung, wurden für das Ergebnisgespräch herangezogen und flossen in die Ergebnisdarstellung ein. Diese gliedert sich wie folgt: Abschnitt 5.1 fasst die Ergebnisse zum Kernbereich 0 des Instruments zusammen. Vor der Darstellung der Analysebefunde zum Feedbackverhalten der beiden MAPT-Apps in Abschnitt 5.3 wird zur besseren Verständlichkeit sowie mit Blick auf die Frage der didaktischen Einbettung des Feedbacks in Abschnitt 5.2 knapp der Ablauf einer typischen Übungssequenz innerhalb der beiden Anwendungen beschrieben. Die Darstellungen des Feedbacks in Abschnitt 5.3 beziehen sich beim *Online Aussprachetrainer* auf das gesamte Angebot[8] und bei *Sylby* zunächst auf die frei zugänglichen Übungen zu segmentalen Phänomenen. Davon ausgehend wird das etwas andere Feedbackformat zu Melodiebewegungen (bei *Sylby* „Intonation“) kontrastierend-ergänzend beschrieben.

## 5. Analyseergebnisse

### 5.1 Kurzbeschreibung der Applikationen

Der *Online Aussprachetrainer* des *Goethe Instituts* kann vollständig kostenfrei über den Browser abgerufen und in den Sprachen Deutsch, Englisch und Vietnamesisch verwendet werden. Eine Registrierung (Name, E-Mail, Land/Region) ermöglicht das Einsehen von Fortschrittsberichten, ist aber für die Nutzung der Anwendung nicht zwingend erforderlich. Zur Nutzung der App müssen in jedem Fall jedoch Angaben zu „Muttersprache“ und Geschlecht (Mann, Frau, Divers) gemacht werden. Die Angabe von Alter und Sprachniveau (A1-C2) ist hingegen fakultativ (vgl. Goethe-Institut o. J.a). In den 12 zur Verfügung stehenden Bereichen (à drei Lektionen) können phonetische Phänomene auf segmentaler (Konsonanten[cluster] und Vokale) und suprasegmentaler (Wortakzent) Ebene geübt werden, wobei einige Übungen die Kombination mehrerer Phänomene in der direkten Abgrenzung zueinander (z. B. ungerundete vs. gerundete Vorderzungen-

[8] Der *Online Aussprachetrainer* wurde offensichtlich in den Wochen vor der Finalisierung dieses Beitrags im August 2024 überarbeitet. Die Ergebnisdarstellung bezieht sich auf diese zuletzt verfügbare Version.

vokale) zulassen. Die der App und Bewertung zugrundeliegende(n) Sprachnorm(en) werden nicht offengelegt, ebenso wenig die Art und Weise, wie die Bewertung der Lernendenspracheingaben erfolgt.

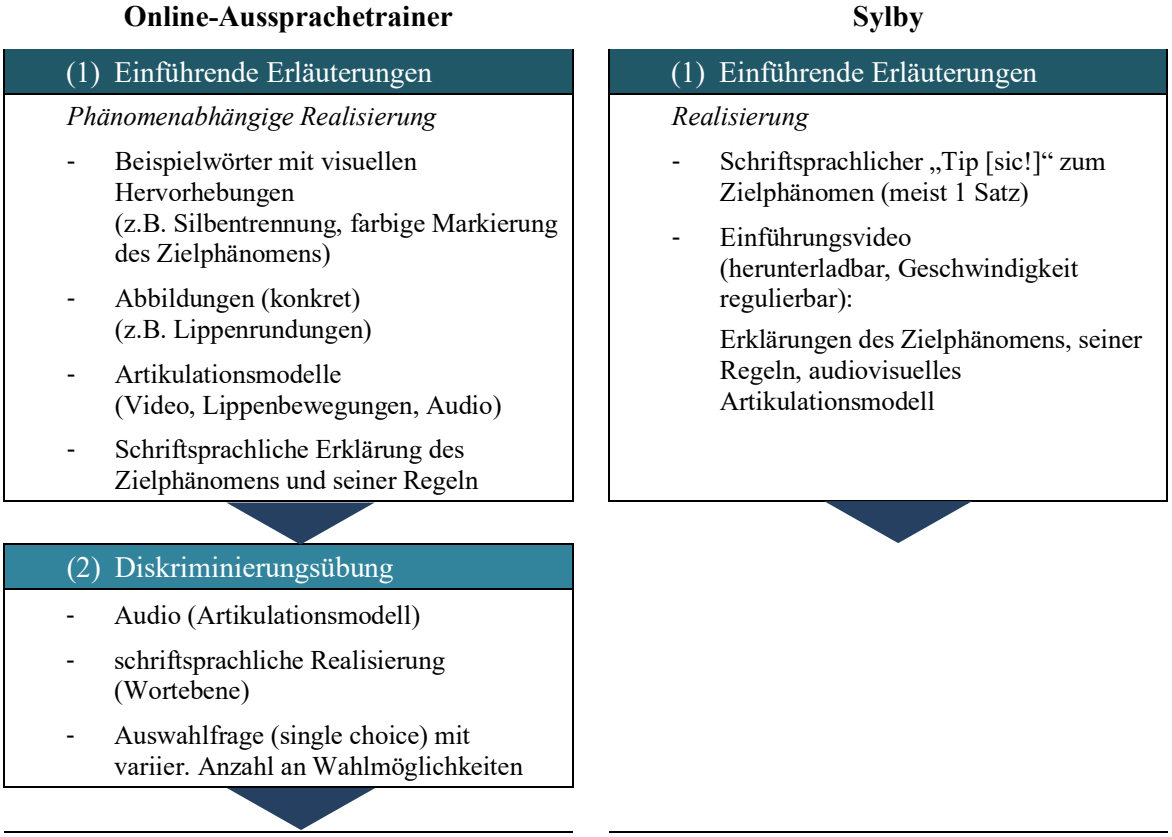
*Sylby* ist eine MAPT-Anwendung der *sylby GmbH*, die in den Sprachen Deutsch und Englisch als App auf iOS- und Android-Endgeräten verwendet werden kann. Die kostenfreie Version enthält nur eine beschränkte Auswahl an Lektionen und Übungen, während die kostenpflichtige Version (Stand Oktober 2024: 12,99€ pro Monat bzw. 69,99€ für 12 Monate) Zugriff auf alle Inhalte gewährt. Name, E-Mail-Adresse und Passwort werden ebenso gespeichert wie Informationen zur „Muttersprache“. Zudem speichert *Sylby* alle Sprachaufnahmen – DSVGO-konform auf einem Server in Deutschland. Hinsichtlich des Feedbackangebots lässt sich folgende Selbstbeschreibung auf der Webseite finden: „Sprechen Sie mit Selbstbewusstsein: Sofortiges KI-Feedback zur Aussprache. Sie müssen nur sprechen und unsere KI wird Ihnen sagen, wie Sie sich verbessern können“ (sylby o. J.). In der kostenfreien Version standen zum Erprobungszeitpunkt 12 von über 60 Einheiten zur Verfügung, in denen eine Auswahl an Segmentalia (Konsonanten[cluster] und Vokale), Suprasegmentalia (Melodiebewegungen, Akzentuierung) und die Kombination mehrerer Phänomene – im Übungsmodus beschränkt auf einzelne Kontrastpaare und ähnliche Phänomene wie zum Beispiel die Lenis-/Fortis-Plosivpaare – geübt werden können. Auch *Sylby* gibt keine Auskunft darüber, welche Sprachnorm(en) dem Algorithmus zugrunde liegen und auf welche Art und Weise eine Bewertung von Lernendenspracheingaben stattfindet. Das Angebot scheint sich zudem in laufender Weiterentwicklung/Erweiterung zu befinden.

### 5.2 Didaktischer Ablauf

Der *Online Aussprachetrainer* des *Goethe Instituts* folgt im Aufbau des Ausspracheтrainings grundsätzlich dem Dreischritt: (1) theoretischer Input / einführende Erläuterungen, (2) Diskriminierungs-/Identifikationsübung, (3) Ausspracheübung. *Sylby* hingegen geht vom (1) theoretischen Input direkt zur (2) Ausspracheübung über (s. Überblick in Abbildung 5).

Die einführenden Erläuterungen enthalten in unterschiedlichem Ausmaß Beschreibungen und Modellierungen des Zielphänomens. Während *Sylby* diese Informationen immer in Form eines einführenden Videos (audiovisuell) realisiert, greift der *Online Aussprachetrainer* je nach Zielphänomen auf unterschiedliche Modi zurück (schrift-

sprachlich, audio, audiovisuell, visuell) und erleichtert so ein gezieltes nochmaliges „Nachschlagen“ der Regeln zu einem späteren Zeitpunkt. Anders als *Sylby* schließt der *Online Aussprachetrainer* den Erläuterungen eine Reihe geschlossener Übungen zur Lautdiskriminierung/-identifikation an, bevor er zur Artikulationsübung überleitet. Diese enthält in beiden Applikationen eine schriftsprachliche Darstellung des Sprechtextes und ein (auditives) Artikulationsmodell, das durch die eigene Aufnahme nachgeahmt werden soll. Im *Online Aussprachetrainer* schließt sich an die Ausspracheübung ein Aussprachetest an, der kein Artikulationsmodell mehr liefert. Bei *Sylby* liegt innerhalb der Ausspracheübungen zudem eine IPA-Transkription vor, welche allerdings durch Unvollständigkeiten (z. B. fehlende Markierung des Wort-/Aussageakzents) und Inkonsistenzen (z.B. <dir> und <wir> unterschiedlich mit/ohne Längung des i-Lautes transkribiert) auffällt und sichtlich keiner der für das Deutsche kodifizierten Aussprachenormen konsequent folgt. Während sich die abspielbare Modellaussprache beim *Online Aussprachetrainer* nur auf Wort- bzw. Wortgruppenebene bewegt, inkludiert *Sylby* teilweise kontextualisierte Beispiele auf Satzebene.



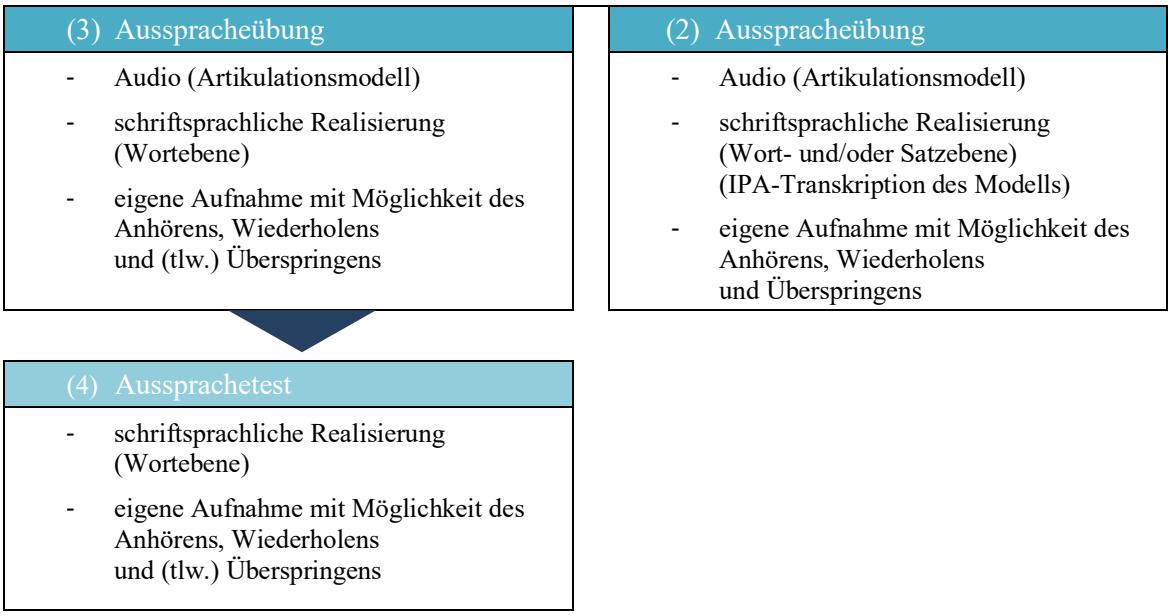



Abb. 5: Didaktische Einbettung – *Online Aussprachetrainer* vs. *Sylby*

### 5.3 Fokus Feedbackverhalten

Das Feedback beider Applikationen gestaltet sich mit Blick auf die Analyse mittels Kriterienraster wie folgt (Detailanalyse in Anhang 2).

**Feedbackverhalten beim *Online Aussprachetrainer***: Der *Online Aussprachetrainer* gibt Nutzenden in seiner Version vom August 2024 für die segmentalen und supra-segmentalen Zielphänomene auf Einzelwortebene multimodales Feedback. Die Bewertung der Aussprache erfolgt nach der Spracheingabe der Lernenden graduell auf visueller sowie schriftsprachlicher Ebene. Je nachdem wie die Applikation die nutzendenseitige Realisierung des Zielsprachenphänomens bewertet, wird das Zielphänomen im Zielwort grün oder rot eingefärbt und gegebenenfalls um weitere Hervorhebungen (z. B. Vokalquantität mittels Punkt oder Strich) ergänzt – zuvor ist es im Zielwort nicht gesondert markiert. Diese Kennzeichnungen können somit auch als unterstützendes Anbieten von qualitativem Wissen nach einem ersten Artikulationsversuch gewertet werden und eine Fokussierung auf den Ziellaut fördern. Unterstützt wird das visuelle Feedback durch ein schriftsprachliches „Richtig!“ (grün), „Das war fast richtig! Versuch es noch einmal!“ (gelb) oder ein „Versuch es noch einmal!“ (rot) sowie eine Umrahmung der Übung in derselben Farbe. Erstere Bewertung, und nur diese, wird zudem mit dem Icon eines grünen Hakens versehen. Durch die mittlere Kategorie (gelb) wird das auf den ersten Blick kategoriale Feedback zu einem graduellen Feedback, wenngleich Nutzenden unklar bleibt, warum die Spracheingabe

seitens des Algorithmus als fast richtig (gelb) im Gegensatz zu falsch (rot) erkannt wurde.

Hinsichtlich des Wissens über das Ziel bietet der *Online Aussprachetrainer* neben vorgelagerten Tipps (s. Abb. 5) in der konkreten Übungs- und Feedbackphase ein auditives Aussprachemodell des einzusprechenden Wortes und nach der Spracheingabe die oben bereits genannten Markierungen zur Fokussierung (und zielsprachlichen Artikulation) des Zielphänomens in der Übung. Die Lernenden entscheiden selbst, ob sie die Modellaussprache abspielen oder ausschließlich auf Basis der ebenfalls angebotenen schriftsprachlichen Sprechvorlage die Spracheingabe vornehmen. Auf eine IPA-Transkription innerhalb der Übungen selbst (siehe dazu die Ergebnisse zu Sylby in Abschnitt 5.3.2) verzichtet der *Online Aussprachetrainer* bei der Darstellung von Wissen zum Zielphänomen. Stattdessen können Nutzende über ein Fragezeichenicon jederzeit auf die einleitenden Informationen zurückgreifen (Pop-Up), ohne die Übung dafür verlassen zu müssen. Allerdings ist dieses (erneut angebotene) Wissen in keiner Weise auf die konkrete Spracheingabe hin adaptiert, sondern stellt das allgemein in der Anwendung zur Verfügung gestellte Wissen dar.

Hilfestellungen zur Erreichung des Zielphänomens bietet die App nach einem ersten (und weiteren) Artikulationsversuch(en) durch das vergleichende Hören der eigenen Spracheingabe mit dem Audiomodell. Explizit aufgefordert werden Lernende zu diesem Vergleich allerdings nicht. Nach einer zweiten Eingabe, die als nicht bzw. nur beinahe korrekt bewertet wird, erhalten die Lernenden zudem qualitatives, gegenstandsbezogenes Feedforward in schriftsprachlicher Form zur Zielerreichung. Hierbei handelt es sich um einen kurzen, phänomenbezogenen Artikulationstipp (z. B. „Sprich <a> mit ungespanntem Mund (= kurz)!“, „Sprich die Silbe langsamer, deutlicher und lauter!“), der sich aus den einführenden Erläuterungen ableiten lässt. Er ist damit keine individualisierte, adaptive Rückmeldung auf eine konkrete Realisierung durch die Nutzenden (z. B. „Dein <a> ist zu lang und zu breit, sprich…“), sondern in Bezug auf die konkrete Übung und das dort fokussierte Zielphänomen vorformuliert.

Ab diesem Zeitpunkt können Lernende entweder weitere Spracheingaben tätigen und bewerten lassen oder aber zur nächsten Aufgabe springen. Dies ist beliebig oft möglich. Die App gestaltet eine Wiederholung der Eingabe gesteuert durch die schriftsprachliche Rückmeldung „Versuch es noch einmal!“ (gelb und rot). Die Möglichkeit dieser Wiederholung existiert dabei jedoch nur so lange, bis man die einzelne Übung über den

„Weiter"- oder „Überspringen"-Button entweder selbstgesteuert verlässt oder durch die App nach ca. drei Versuchen dazu gezwungen wird: Das Mikrophon wird in diesem Fall grau und kann nicht mehr angesteuert werden. Eine Rückkehr zu einer Einzelübung in Form eines „Zurück"-Buttons ist zu keinem Zeitpunkt möglich.

Zusätzlich zu dem graduellen Feedback zum Lernstand innerhalb der einzelnen Übungen erhalten die Nutzenden am Ende einer Übungsabfolge ein summatives Feedback, das aus einer zusammenfassenden Übersicht aller Zielwörter, dem erreichten Score sowie aus der Anzahl der Versuche für die einzelnen Zielwörter besteht. Die Zusammenfassung („Geschafft! X von X Übungen") bietet auf diese Weise ein summatives Feedback, das nutzendenseitig Rückschlüsse darüber zulässt, welche Zielwörter bezogen auf das Zielphänomen Schwierigkeiten bereitet haben. Konkrete Handlungsanweisungen oder Tipps, wie eine weitere Verbesserung des Scores bzw. der Aussprache des Zielphänomens erreicht werden kann, werden hier nicht gegeben.

Die Erprobung der Anwendung zeigt zudem, dass die Qualität der Spracherkennung sowohl zwischen einzelnen Beispielwörtern zu einem Zielphänomen als auch zwischen den Zielphänomenen stark schwankt und die Bewertungen der Eingabe teilweise kritisch zu beurteilen sind: So konnte mehrfach beobachtet werden, dass das Einsprechen deutlich abweichender Laute eine positive Bewertung durch die Spracherkennung erfuhr, während als korrekt intendierte Realisierungen auch nach wiederholter Aufnahme und Eingabe weiterhin als falsch gewertet wurden. Zudem schwankt die Funktionalität des Mikrophons, das den Beginn der Aufnahme durch einen Countdown (3-2-1) anzeigt: Teilweise werden Spracheingaben überhaupt nicht erkannt oder das pulsierende Mikrophon reagiert stark verzögert.

**Feedbackverhalten bei *Sylby***: Bei *Sylby* erhält man für die segmentalen Zielphänomene ebenfalls multimodales Feedback zum eigenen Lernstand. So erfolgt eine farbliche Fettmarkierung (rot/grün) inkl. Unterstreichung sowohl in der IPA-Transkription als auch in der schriftsprachlichen „Sprechvorlage" für die Übungsitems. Zudem erhält man eine schriftsprachliche Rückmeldung: Bei dieser werden keine kategorialen Ausdrücke wie „richtig" oder „falsch" verwendet, sondern die Aussagen „Perfekt! Du hast ein Händchen für Sprache" (bei einer durch die App richtig bewerteten Aussprache des Zielphänomens) und „Noch nicht ganz gelungen. Versuch es doch nochmal" (bei einer falsch bewerteten Aussprache des Zielphänomens). Trotz dieses Wortlautes sind diese Rückmeldungen als kategorial zu werten, da in der Durchführung mehrerer

Übungen keine anderen Aussagen ausgemacht werden konnten, die darauf hindeuten, dass diese Rückmeldungen graduell über mehr als zwei Stufen je nach Qualität der Sprecheingabe variieren. Zusätzlich wird das jeweilige schriftsprachliche Feedback visuell kategorial durch die Farben grün und rot unterstützt. Das Feedback ist damit primär kategorial gestaltet, gegenstandsbezogen aber dahingehend, dass farblich genau ausgewiesen wird, welche Zielphänomene in der Spracheingabe als falsch (rot) oder richtig (grün) gewertet wurden – es also zu qualitativen Hervorhebungen in der schriftsprachlich vorliegenden Sprechvorlage (gefettet, eingefärbt und unterstrichen) kommt. Graduell gestaltet ist das Feedback am Ende jeder Übungsabfolge zu einem bestimmten Zielphänomen. Hier wird der Anteil der richtig gewerteten Spracheingaben errechnet und als Prozentzahl ausgegeben. Es handelt sich dabei aber um ein summatives Feedback, welches im Gegensatz zum *Online Aussprachetrainer* keinerlei Ausdifferenzierung nach Übungsitems anbietet und auch keine unmittelbare formative Hilfestellung für weitere Artikulationsversuche liefert.

Zudem zeigte sich beim Feedbackverhalten der App in der Erprobung, dass nur das Zielphänomen für die Bewertung als richtig oder falsch herangezogen wird. Die Artikulation des Lautkontextes in den Übungswörtern bzw. -sätzen wurde für die Generierung des Feedbacks nicht berücksichtigt. Die Qualität der zugrundeliegenden Spracherkennung, auf der die Bewertung basiert, scheint – so der Eindruck in den Erprobungen – dabei wie beim *Online Aussprachetrainer* zwischen den einzelnen Zielphänomenen zu variieren[9]. Ebenso kann ergänzend angemerkt werden, dass das Feedback auf die Spracheingabe wie auch beim *Online Aussprachetrainer* unmittelbar erfolgt. Man hat als lernende Person also in beiden Applikationen nicht die Steuerungsmöglichkeit, die Spracheingabe mehrfach zu versuchen, bevor man diese durch die App bewerten lässt. Dafür gestaltet sich die Spracheingabe bei *Sylby* selbst als funktionaler, da sie sich gezielt starten und stoppen lässt und in den Erprobungen deutlich weniger Fehlfunktionen aufwies.

Bezüglich des Wissens über das Zielphänomen verfolgt *Sylby* nicht nur durch die vorgelagerten Erklärvideos, sondern auch direkt durch die Gestaltung der Übungs- und Feedbackphase das Prinzip, entsprechendes Wissen vor und nach der Spracheingabe

[9] Wird beispielsweise ein abweichendes Wort (z. B. „Hut“) eingesprochen, das den Hauchlaut als Ziellaut enthält, aber nicht dem Zielwort (z. B. „herrlich“) entspricht, wertet dies der Algorithmus als richtige Eingabe. Er scheint damit Einzellaute in seiner Bewertung zu fokussieren.

bereitzustellen. So liegt neben einer Verschriftlichung der einzusprechenden Sätze und Wörter die (wenn auch hinsichtlich ihres Beschreibungscharakters abstrakte) IPA-Transkription der einzusprechenden Wörter vor. Ebenso ist es jederzeit möglich, mittels eines Pfeilicons zum Erklärvideo zurückzukehren. Darüber hinaus wird ein auditives Aussprachemodell angeboten, welches vor und nach der eigenen Spracheingabe (mehrfach) angehört[10] werden kann. Die App ist dabei so konfiguriert, dass man beim Aufrufen einer neuen Übung das Aussprachemodell zum einzusprechenden Wort oder Satz ohne weitere Manipulation bereits ein erstes Mal hört. Eine Abweichung von diesem Prinzip ist nur dadurch erkennbar, dass die Entsprechungen der Ziellaute in der schriftsprachlichen Sprechvorlage sowie in der IPA-Transkription erst nach einem ersten Einsprechversuch durch farbliche Fettmarkierung (rot/grün) inkl. Unterstreichung hervorgehoben werden. Eine Kennzeichnung der Zielphänomene im Sinne einer Aufmerksamkeitslenkung vor der ersten Spracheingabe wird nicht angeboten.

Hinsichtlich der Wege zum Ziel fallen die Angebote der App, im Vergleich zum *Online Aussprachetrainer*, eher knapp aus. Gegeben ist aber ebenfalls die Möglichkeit eines vergleichenden Hörens. So kann man sich nach einer ersten Spracheingabe sowohl das Aussprachemodell als auch die eigene lerner:innensprachliche Version anhören. Ebenso wird eine Bewertung als „falsch“ mit der Empfehlung („Versuch es doch nochmal“) und der Möglichkeit verknüpft, eine erneute Spracheingabe durchzuführen. Ein Vor- und Zurückgehen zwischen einzelnen Übungen innerhalb einer Übungskette ist dabei im Gegensatz zum *Online Aussprachetrainer* jederzeit möglich.

Abschließend ist zu ergänzen, dass für das prosodische Zielphänomen der „Intonation“ (Melodieverläufe) das Feedback in *Sylby* anders gestaltet ist. So wird hier keine IPA-Transkription angeboten, stattdessen vor und nach der Spracheingabe eine stark schematisierte Darstellung der Zielmelodie (nämlich als schräg ansteigende/absinkende Gerade), darüber hinaus wie bei den segmentalen Phänomenen eine schriftsprachliche Sprechvorlage (inkl. Interpunktionszeichen wie Frage-/Ausrufezeichen) sowie ein auditives Aussprachemodell. Nach der Spracheingabe wird der lerner:innensprachliche Melodieverlauf (unabhängig von der Qualität der Realisierung) grün im selben Bild wie die schematisierte Zielmelodiebewegung (gelb) dargestellt. Diesem scheint eine f0-

[10] Technisch ist es Nutzenden dabei sowohl bei *Sylby* als auch beim *Online Aussprachetrainer* möglich, das Modellaudio abzuspielen und gleichzeitig den Aufnahmeknopf zu drücken, sodass letztendlich das Modell und nicht eine eigene Aufnahme bewertet würde.

Analyse zugrunde zu liegen, da für die Darstellung des Melodieverlaufs der Spracheingabe keine Gerade dargeboten, sondern ein Linienverlauf ersichtlich wird. Da allerdings keine weitere Rückmeldung zur lerner:innensprachlichen Eingabe erfolgt – auch ein Anhören der eigenen Spracheingabe ist an dieser Stelle nicht möglich – und somit immer eine Abweichung zwischen stark abstrahierter und vereinfachter Zielmelodiedarstellung (Gerade) und der stärker am Grundfrequenzverlauf orientierten Visualisierung der lerner:innenseitigen Spracheingabe (Linienverlauf) gegeben ist, bleibt die Verständlichkeit dieser Feedbackform fragwürdig. Das abschließende (für die anderen Zielphänomene summative) Feedback der Übungsabfolge besteht in diesem Fall lediglich aus einem „Geschafft! Bleib dran und nimm die nächste Übung gleich in Angriff".

## 6. Implikationen

Aus der Erprobung und Analyse beider Anwendungen lassen sich folgende allgemeine Implikationen für eine (Weiter-)Entwicklung der Applikationen und eine damit einhergehende Forschung ableiten:

**Feedback:** Hinsichtlich des Feedbackverhaltens zeigen sich durchaus mehrere positive Aspekte bei beiden oder einer der beiden MAPT-Apps, so zum Beispiel die Multimodalität in den Darstellungsformen, die Möglichkeit zum vergleichenden Hören und die (wenn auch unterschiedlich stark gegebenen) Freiheitsgrade in der eigenständigen Navigation durch die Übungsangebote, die zum Beispiel auch eine Rückkehr zu den bzw. ein Aufrufen der einleitenden Erklärungen und weiteren didaktischen Angebote zu den Zielphänomenen ermöglichen. Kritisch zu sehen ist zugleich aber, dass die Rückmeldungen primär kategorial in einer Dichotomie von richtig/falsch oder graduell mit einer Zwischenstufe (fast richtig) konzipiert sind und damit vor allem das Potenzial gegenstandsbezogener Rückmeldungen zum Lernstand nur ansatzweise nutzen. Die Version des *Online Aussprachetrainers* vom August 2024 zeigt allerdings eine positive Entwicklung in Form einer stärkeren Einbettung qualitativer Angebote zur Zielerreichung. Schuldig bleiben die beiden MAPT-Apps aber nach wie vor eine auf die Sprecheingabe hin adaptierte Darstellung von Wissen zur Zielerreichung, sodass sich Kelz' These zur Bedeutung der Lehrperson für den Aussprachetrainer aus dem Jahr 1992 auf Basis der hier erzielten Analyseergebnisse mehr als 30 Jahre später (noch) nicht widerlegen lässt.

**Phonetische Kontrastivität:** Obwohl beide Anwendungen Erstsprachen („Muttersprachen“) erfragen, spielen diese bei der Darstellung oder Fokussierung bestimmter phonetischer Phänomene keine Rolle. Auch die Angabe eines Sprachniveaus (*Online Aussprachetrainer*) wirkt sich nicht auf die Präsentation der Inhalte (Reihenfolge, Auswahl an Phänomenen, etc.) oder Bewertung (Toleranz, Feedbackverhalten) durch die Anwendungen aus. Diese Art der Datenerfassung dient damit keinen differenzierenden gegebenenfalls sprachkontrastiven, sondern vielmehr rein statistischen Zwecken.

**Diversität und Standardism**: Die Modellsprechenden in beiden Anwendungen zeigen wenig Diversität mit Blick auf Geschlecht und Alter. Sprachliche Diversität wird nicht realisiert: Es findet eine Orientierung an einem bundesdeutschen Standard statt, während sprachliche Varietäten des Deutschen kaum[11] bis keine Berücksichtigung finden – auch nicht in Aussprache-Beispielen. Eine Omnipräsenz der bundesdeutschen Aussprachevariante bleibt aber auch abseits der untersuchten MAPT-Apps Realität: Weder die im deutschsprachigen Raum meistgenutzten MT-Tools *Google Translate* oder *DeepL* noch *Microsoft Word* lassen bislang eine Auswahl an Standardvarietäten bei Eingabe (Mikrofon) oder Ausgabe (Vorlesefunktion) zu (Stand November 2024).[12]

**Transparenz**: Beide Anwendungen machen keine Angaben darüber, auf welcher Datengrundlage der ASR-Algorithmus trainiert wurde, wie er zu seinen Berechnungen kommt und auf welcher Basis damit Lernendensprache bewertet wird. Letzteres bleibt ein mathematisches Unterfangen und bildet statistische Wahrscheinlichkeiten ab, die sich entsprechend im Feedbackverhalten der beiden Tools – richtig/teilweise richtig/falsch, prozentuale „Richtigkeit“ – niederschlagen. Die berechnete, aber nicht offengelegte Wahrscheinlichkeit sagt jedoch nichts darüber aus, ob die Nutzenden einen oder mehrere Laute (technisch) „korrekt“ realisiert haben. Auch bleibt unklar, ob die Äußerung in einer interpersonalen Kommunikationssituation verstanden würde. Für

[11] Bei *Sylby* lässt sich im Erklärvideo zum Laut [x] ein Verweis auf schweizerdeutsche Realisierungen finden sowie eine Selbstpositionierung der Sprecherin für das „Standarddeutsche“: „Also wenn du Probleme mit dem Laut [x] hast, kannst du immer noch in die Schweiz ziehen. Wenn du aber lieber Berlin als Bern magst, solltest du hier bleiben“ (Minute 0:44-0:55).

[12] Andere KI-Anwendungen, die keine expliziten MAPT-Apps darstellen, ermöglichen eine Auswahl von Geschlecht und Stimmfarbe des KI-Modellsprechers (z.B. *elevenlabs*), deutscher Standardvarietäten für ein Gespräch mit einem Avatar (z.B. *Gliglish*) oder den Wechsel in Varietäten und auch Dialekte durch gezieltes Prompten (z.B. *ChatGPT Advanced Voice Mode*). Die Qualität der Realisierungen auf segmentaler und suprasegmentaler Ebene schwankt hier je nach Anwendung.

eine Folgestudie denkbar wäre es daher, die jeweiligen Spracheingaben parallel auch zu autographieren und einer akustischen Analyse zu unterziehen. Dadurch könnten die Spannweiten an möglichen Realisierungen, die das Tool als korrekt wertet, ermittelt und Aussagen über seine Akkuratheit getroffen werden.

**Korrektheit:** Die Qualität bzw. fachlich-inhaltliche Korrektheit der didaktischen Angebote in den Apps stand nicht im Zentrum der Analyse und wurde daher auch nur punktuell (z. B. hinsichtlich der IPA-Transkription bei *Sylby*) angesprochen. Diese peripheren Thematisierungen zeigen aber bereits einzelne kritische Punkte (z. B. hinsichtlich der Präzision, der Passung der Anweisungen und Erklärungen mit den konkreten Übungsitems sowie der Korrektheit bei der IPA-Transkription) auf, die in einer weiteren Detailanalyse einer genaueren Überprüfung unterzogen werden könnten.

## 7. Fazit

Wie bereits durch die Identifizierung möglicher MAPT-Apps gezeigt, fokussieren die meisten Apps ein Aussprachetraining mit Blick auf die englische Sprache und ihre Varietäten, während sich für das gezielte Aussprachetraining im Kontext DaF/DaZ grundsätzlich nur wenige Anwendungen finden lassen, vor allem dann, wenn eine kostenfreie Nutzung als Auswahlkriterium angesetzt wird. Die wenigen vorhandenen Applikationen wiederum machen bislang nicht von allen – bereits bei Kaiser (2018) und Walesiak (2017) erwähnten – technischen Möglichkeiten in Bezug auf Feedback Gebrauch. So enthielten die hier analysierten Anwendungen zum Beispiel keinerlei Formen videobasierten Feedbacks. Während einführende Erklärungen, Visualisierungen und Beispiele grundsätzlich als positiv und informativ bewertet werden können (Wissen über das Ziel), zeigten beide Apps in unseren Erprobungen gerade auf der Ebene des Feedbacks (Wissen über Lernstand) und dessen technischer Umsetzung Verbesserungspotenzial. Neben der Handhabung der Aufnahmefunktion (starten, stoppen, wiederholen) erwies sich vor allem die Bewertung der eigenen Spracheingaben durch die Anwendungen als problematisch: Da die Bewertungsgrundlage selbst nicht transparent gemacht wird und teilweise Fragen ob ihrer Korrektheit und Toleranzwerte aufwarf, bot das sich an sie anschließende kategoriale oder graduelle und nur ansatzweise gegenstandbezogene Feedback kaum passgenaue Hilfestellung zur (weiteren) Verbesserung der Aussprache der Zielphänomene. Mittels vergleichendem Hören, einem Rückgriff auf allgemeine Regeln und Erläuterungen sowie Übungs- bzw. Aufnahmewieder-

holungen ist es in den allermeisten Fällen an den Nutzenden, sich Wissen über die Wege zum Ziel selbst zu erschließen. Der *Online Aussprachetrainer* bietet hier zwar teilweise durch phänomenbezogene Aussprachetipps Abhilfe, jedoch nur nach mehrfacher Fehleingabe und ohne direkten Bezug zur lernendenenseitigen Spracheingabe. Ein positives Feedback zu einer gelungenen Realisierung erfolgt in beiden Apps rein kategorial.

Um einen Einsatz von MAPT-Apps im DaF/DaZ-Unterricht zukünftig zu unterstützen, bräuchte es demnach auf Seiten der Anwendungen

- Transparenz in Bezug auf Datengrundlage, Training und die an der Entwicklung beteiligten Personen,
- Transparenz in Bezug auf Bewertungsverfahren und -grundlagen (Sprachnormen, Statistik) sowie
- ein diversifiziertes, multimodales Feedback, das Lernenden Rückschlüsse dahingehend erlaubt,
  a) wie die KI ihre Aussprache bewertet hat,
  b) in welcher Weise sich ihre Sprecheingabe tatsächlich von dem zugrundeliegenden Aussprachemodell unterscheidet und
  c) wie sie sich konkret ausgehend vom individuellen Lernstand, den ihre Spracheingabe darstellt, verbessern können.

Vor allem der letzte Punkt c) beruht stark auf dem Adaptieren und Anpassen von Feedback auf die konkrete lerner:innenseitige Artikulation, sodass sich auch gerade in diesem Punkt die von Kelz (1992) postulierte Alleinstellung der Lehrperson im Ausspracheunterricht auf Basis der in diesem Beitrag gezeigten Ergebnisse zu zwei MAPT-Apps für DaF/DaZ nach wie vor nicht widerlegen lässt. Es bleibt zu beobachten, wie sich KI-gestützte Angebote in ihrer Adaptivität in den nächsten Jahren weiterentwickeln werden – vor allem wann und wie die durch viele Anwendungen angepriesenen Möglichkeiten einer Binnendifferenzierung, eines individualisierten Lernpfades und individualisierten Feedbacks tatsächlich eine Umsetzung erfahren.

## Bibliographie

Ashwell, Tim; Elam, Jesse R. (2017) How Accurately Can the Google Web Speech API Recognize and Transcribe Japanese L2 English Learners' Oral Production?. *Jalt CALL Journal* 13 (1), 59-76. DOI: https://doi.org/10.29140/jaltcall.v13n1.212 .

Ata, Murat; Debreli, Emre (2021) Machine translation in the language classroom: Turkish EFL learners' and instructors' perceptions and use. *IAFOR Journal of Education* 9 (4), 103-122. DOI: https://doi.org/10.22492/ije.9.4.06.

Bin Dahmash, Nada (2020) I can't live without Google Translate: A close look at the use of Google Translate App by second language learners in Saudi Arabia. *Arab World English Journal* 11 (3), 226-240. DOI: https://doi.org/10.24093/awej/vol11no3.14.

Bliss, Heather; Abel, Jennifer; Gick, Bryan (2018) Computer-Assisted Visual Articulation Feedback in L2 Pronunciation Instruction: A Review. *Journal of Second Language Pronunciation* 4 (1), 129-153.

Cardoso, Walcir (2018): „Learning L2 Pronunciation with a Text-to-Speech Synthesizer". In: *Future-Proof CALL: Language Learning as Exploration and Encounters – Short Papers from EUROCALL 2018*, 16–21. DOI: https://doi.org/10.14705/rpnet.2018.26.806.

Cucchiarini, Catia; Neri, Ambra; De Wet, Febe; Strik, Helmer (2007) ASR-Based Pronunciation Training: Scoring Accuracy and Pedagogical Effectiveness of a System for Dutch L2 Learners. *Interspeech 2007*, 2181-2184. DOI: https://doi.org/10.21437/Interspeech.2007-594.

Dieling, Helga; Hirschfeld, Ursula (2000) *Phonetik lehren und lernen*. Stuttgart: Klett Sprachen.

Dlaska, Andrea; Krekeler, Christian (2008) Self-Assessment of Pronunciation. *System* 36, 506-516.

Dlaska, Andrea; Krekeler, Christian (2013) The Short-Term Effects of Individual Corrective Feedback on L2 Pronunciation. *System* 41, 25-37.

Fouz-González, Jonás (2020) Using Apps for Pronunciation Training: An Empirical Evaluation of the English File Pronunciation App. *Language Learning & Technology* 24 (1), 62-85.

Goethe-Institut (o. J.a) *Goethe Aussprachetrainer*. Online: https://aussprachetraining.goethe.de/ (13.6.2024).

Goethe-Institut (o. J.b) *Online Aussprachetrainer - Goethe-Institut*. Online: https://www.goethe.de/de/spr/ueb/ast.html (28.2.2024).

He, Yue; Cardoso, Walcir (2021) Can Online Translators and Their Speech Capabilities Help English Learners Improve Their Pronunciation?. *CALL and Professionalisation: Short Papers from EUROCALL 2021*, 126-131. DOI: https://doi.org/10.14705/rpnet.2021.54.1320.

Hirschfeld, Ursula; Reinke, Kerstin (2016) *Phonetik im Fach Deutsch als Fremd- und Zweitsprache*. Berlin: Erich Schmidt Verlag.

Hirschi, Kevin; Kang, Okim; Hansen, John; Cucchiarini, Catia; Strik, Helmer (2022) Mobile-Assisted Pronunciation Training With Adult Esol Learners: Background, Acceptance, Effort, and Accuracy. *Virtual PSLLT*. DOI: https://doi.org/10.31274/psllt.13272.

Huang, Yuhui; Lee, Andrew H.; Ballinger, Susan (2023) The Characteristics and Effects of Peer Feedback on Second Language Pronunciation. *Journal of Second Language Pronunciation* 9 (1), 47-70.

Jenkins, Jennifer (2000) *The phonology of English as an international language*. Oxford: Oxford University Press.

Jin, Li; Deifell, Elizabeth (2013) Foreign language learners' use and perception of online dictionaries: A survey study. *MERLOT Journal of Online Learning and Teaching* 9 (4), 515-33.

Kaiser, DJ (2018) Mobile-assisted pronunciation training: The iPhone pronunciation app project. *IATEFL Pronunciation Special Interest Group Journal* 58, 38-52 (Preprint 1-11).

Kelz, Heinrich P. (1992) Lernziel deutsche Aussprache. *Materialien Deutsch als Fremdsprache* 32, 23-38.

Kennedy, Olivia (2021) Independent Learner Strategies to Improve Second Language Academic Writing in an Online Course. *CALL and Professionalisation: Short Papers from EUROCALL 2021*, 184-188. DOI: https://doi.org/10.14705/rpnet.2021.54.1330.

Lan, En-Minh (2022) A Comparative Study of Computer and Mobile-Assisted Pronunciation Training: The Case of University Students in Taiwan. *Education and Information Technologies* 27 (2), 1559-1583. DOI: https://doi.org/10.1007/s10639-021-10647-4.

Lee, Junkyu; Jang, Juhyun; Plonsky, Luke (2015) The Effectiveness of Second Language Pronunciation Instruction: A Meta-Analysis. *Applied Linguistics* 36 (3), 345-366.

Lee, Yunhyun (2021) Effectiveness of Mobile Assisted Pronunciation Training in the Acquisition of English Vowels, /o/ and /ɔ/. *Foreign Languages Education* 28 (3), 53-76. DOI: https://doi.org/10.15334/FLE.2021.28.3.53.

Martin, Ines A.; Sippel, Lieselotte (2023) Long-Term Effects of Peer and Teacher Feedback on L2 Pronunciation. *Journal of Second Language Pronunciation* 9 (1), 20-46.

McCrocklin, Shannon (2019a) ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation* 5 (1), 98-118. DOI: https://doi.org/10.1075/jslp.16034.mcc.

McCrocklin, Shannon (2019b) Learners' Feedback Regarding ASR-based Dictation Practice for Pronunciation Learning. *CALICO* 36 (2), 119-137. DOI: https://doi.org/10.1558/cj.34738.

Medvedev, Gennady (2016) Google Translate in Teaching English. *The Journal of Teaching English for Specific and Academic Purposes* 4 (1), 181-193.

Meisarah, Fitria (2020) Mobile-Assisted Pronunciation Training: The Google Play Pronunciation and Phonetics Application. *Script Journal: Journal of Linguistics and English Teaching* 5 (2), 70-88. DOI: https://doi.org/10.24903/sj.v5i2.487.

Moorkens, Joss (2022) Ethics and Machine Translation. In: Kenny, Dorothee (Hrsg.) *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence*. Berlin: Language Science Press, 121-140.

Neri, Ambra; Cucchiarini, Catia; Strik, Helmer; Boves, Lou (2002) The Pedagogy-Technology Interface in Computer Assisted Pronunciation Training. *Computer Assisted Language Learning* 15 (5), 441-467. DOI: https://doi.org/10.1076/call.15.5.441.13473.

Neri, Ambra; Cucchiarini, Catia; Strik, Wilhelmus (2003) Automatic Speech Recognition for Second Language Learning: How and Why It Actually Works. *Speech Communication*, 1157-1160.

Niño, Ana (2020) Exploring the use of online machine translation for independent language learning. *Research in Learning Technology* 28. DOI: https://doi.org/10.25304/rlt.v28.2402.

Niño, Ana (2025) Pronunciation Practice with Google Translate TTS and ASR Functions. *CALICO Journal* 42 (1), 69-93. DOI: https://doi.org/10.3138/calico-2024-1223.

Organ, Alison (2023) Attitudes to the Use of Google Translate for L2 Production: Analysis of Chatroom Discussions among UK Secondary School Students. *The Language Learning Journal* 51 (3), 328-343. DOI: https://doi.org/10.1080/09571736.2021.2023896.

Papin, Kevin; Cardoso, Walcir (2022) Pronunciation Practice in Google Translate: Focus on French Liaison. *Intelligent CALL, Granular Systems and Learner Data: Short Papers from EUROCALL 2022*, 322-327. DOI: https://doi.org/10.14705/rpnet.2022.61.1478.

Papin, Kevin; Cardoso, Walcir (2023) Using Google Translate's Speech Features for Self-Regulated French Pronunciation Practice. *EuroCALL 2023. CALL for All Languages – Short Papers*, 48-53. DOI: https://doi.org/10.4995/EuroCALL2023.2023.17000.

Papin, Kevin; Cardoso, Walcir (2025) Assessing the Pedagogical Potential of Google Translate's Speech Capabilities: Focus on French Pronunciation. *CALICO Journal* 42 (1), 1-24. DOI: https://doi.org/10.3138/calico-2025-0117.

Pennington, Martha C.; Rogerson-Revell, Pamela (2019) *English pronunciation teaching and research. Contemporary perspectives*. London: Palgrave Macmillan.

Rangsarittikun, Ronnakrit (2022) Jumping on the Bandwagon: Thai Students' Perceptions and Practices of Implementing Google Translate in Their EFL Classrooms. *English Teaching & Learning*. DOI: https://doi.org/10.1007/s42321-022-00126-5.

Ross, Rae K.; Lake, Vickie E.; Beisly, Amber H. (2021) Preservice Teachers' Use of a Translation App with Dual Language Learners. *Journal of Digital Learning in Teacher Education* 37 (2), 86-98. DOI: https://doi.org/10.1080/21532974.2020.1800536.

Saito, Kazuya; Lyster, Roy (2012) Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of /ɹ/ by Japanese Learners of English. *Language Learning* 62 (2), 595-633.

Shadiev, Rustam; Sun, Ai; Huang, Yueh-Min (2019) A Study of the Facilitation of Cross-cultural Understanding and Intercultural Sensitivity Using Speech-enabled Language Translation Technology. *British Journal of Educational Technology* 50 (3), 1415-1433. DOI: https://doi.org/10.1111/bjet.12648.

Sylby (o. J) *Speak with confidence – THE AI-POWERED PRONUNCIATION AND LANGUAGE LEARNING APP*. Online: https://sylby.com/de/ (28.02.2024).

Tan, April (2021) Study Intonation: A mobile-assisted pronunciation training application. *The Electronic Journal for English as a Second Language* 25 (3), 1-8.

Walesiak, Beata (2017) Mobile Pron. Apps – a Personal Investigation. *Peak Out! Journal of the IATEFL Pronunciation Special Interest Group* 57, 16-28.

Wirantaka, Andi; Fijanah, Mahdiana Syahri (2021) Effective use of Google Translate in writing. *Advances in Social Science, Education and Humanities Research* 626, 15-23.

## Schlagwörter

Aussprache, Künstliche Intelligenz, DaF-Unterricht, Feedback

## Biographische Angaben

**Katrin Engelmayr-Hofmann** ist Universitätsassistentin (prae-doc) am Fachbereich Deutsch als Fremd- und Zweitsprache (Institut für Germanistik) der Universität Wien. Ihr Forschungsinteresse gilt dem kritisch-reflektierten Einsatz von Technologien im Bereich DaF, betreffend unter anderem die Entwicklung bzw. Vermittlung digitaler Kompetenzen im Umgang mit sogenannter Künstlicher Intelligenz (insbesondere Machine Translation Tools bzw. MT-Literacy) sowohl lehrenden- als auch lernendenseitig.

**Sandra Reitbrecht** ist Vertragshochschullehrperson am Institut für Urban Diversity Education der Pädagogischen Hochschule Wien und Projektmitarbeiterin an der Universität Graz. Sie beschäftigt sich mit Fragestellungen der Aussprachedidaktik im DaF/DaZ-Unterricht sowie der (fortgeschrittenen) Literalitätsentwicklung in sprachlich heterogenen Lernendengruppen.

# Anhang 1

## Kriterienraster zur Analyse von Feedback mit Schwerpunkt Spracheingabe

**0 Allgemeines**

Name: Hersteller:in:

Zugang: ✓ ✗ iOs ✓ ✗ Android ✓ ✗ browserbasiert

Auszüge Selbstbeschreibung: ZITATE der Selbstauskunft durch das jeweilige Tool

✓ ✗ Offenlegung zugrundeliegender Sprachnormen

✓ ✗ Offenlegung der Art/Weise der Bewertung von Lerner:innenspracheingaben

Folgende phonetische Phänomene können geübt werden:

✓ ✗ Segmentalia (Konsonanten, Vokale, etc.) in jeweils einzelnen Übungen

✓ ✗ Suprasegmentalia (Wortakzent, Melodiebewegungen, Rhythmus, etc.) in jeweils einzelnen Übungen

✓ ✗ Kombination mehrerer Phänomene in einer Übung

**1 Wissen über Lernstand**[13]

Welche Art(en) von Rückmeldung(en) zum Lernstand erfolgen über die App?

| **Modus?** | **Art?** | **Beispiel?** | **Quellen** |
|---|---|---|---|
| schriftsprachlich | kategorial | *richtig, falsch* | Huang/Lee/Ballinger (2023)<br>*Kaiser (2018)* |
| | graduell | *Prozentzahlen, „Nah dran!“, „Fast!“* | |
| | qualitativ | *gegenstandsbezogen:*<br>*„Dein U klingt wie ein I.“* | Huang/Lee/Ballinger (2023) |
| visuell | kategorial/ graduell | *Farben, Icons* | |
| | qualitativ | *Hervorhebungen im Text (fett, kursiv, Silben, Rhythmus, Textanimation)* | Hirschi et al. (2022) |
| | | *Spektrogramm* | Bliss/Abel/Gick (2018)<br>*Kaiser (2018)* |
| | | *Melodiekurve (f0)* | Bliss/Abel/Gick (2018)<br>Tan (2021) |
| | | *visuelle Verortung in einer Lauttabelle (formant chart)* | Bliss/Abel/Gick (2018)<br>*Kaiser (2018)* |
| auditiv | kategorial/ graduell | *Soundeffekte* | |

[13] In Anlehnung an die drei Feedback-Schritte bei Dlaska und Krekeler (2013).

**2 Wissen über Ziel**

Wie wird das phonetische Zielphänomen in der Ausspracheübung dargestellt?

| **Modus?** | **Beispiel?** | **… der Spracheingabe?** | **Quellen** |
|---|---|---|---|
| schriftsprachlich | Transkriptionszeichen (IPA-Symbole) | VOR / NACH | Huang/Lee/Ballinger (2023) |
| | Beschreibung des phonetischen Zielphänomens | VOR / NACH | |
| visuell | Hervorhebungen im Text *(fett, kursiv, Silben, Rhythmus, Textanimation)* | VOR / NACH | |
| | Markierungen *(Wellen, Pfeile für Melodiebewegungen etc.)* | VOR / NACH | |
| | Abbildungen (konkret) *(Bild von Lippenrundung etc.)* | VOR / NACH | |
| | Spektrogramm (abstrakt) | VOR / NACH | |
| audiovisuell | Artikulationsmodell (Video) *(Art des Videos, Repräsentation von Diversität (Alter/Geschlecht Sprecher:innen, Registervariationen, phono-stilistische Variationen))* | VOR / NACH | |
| auditiv | rein auditives Aussprachemodell (Audio) *(Repräsentation von Diversität (Alter/Geschlecht Sprecher:innen, Registervariationen, phono-stilistische Variationen))* | VOR / NACH | |

**3 Wissen über Wege zum Ziel**[14]

Welche Arten von Unterstützung zur Aussprachеverbesserung werden gegeben?

| **Modus?** | **Beispiel?** | **Quellen** |
|---|---|---|
| implizit | vergleichendes Sehen *(eigene Videoaufnahme vs. Modellvideo)* *(eigenes Spektrogramm vs. Zielspektrogramm)* | Saito/Lyster (2012) |
| | vergleichendes Hören *(eigene Audioaufnahme vs. Modellaudio)* | Dlaska/Krekeler (2013) Saito/Lyster (2012) |
| implizit adaptiv | (erneutes) Artikulationsmodell (Video) *(mit Überbetonung, Verlangsamung, unterstützenden Gesten etc.)* | Bliss/Abel/Gick (2018) |
| | (erneutes) rein auditives Aussprachemodell (Audio) *(mit Überbetonung, Verlangsamung etc.)* | Huang/Lee/Ballinger (2023) |
| explizit | Erläuterungen/Anweisungen (schriftlich, visuell, auditiv), um Ziel zu erreichen *(Ableitung von anderen Lauten/Geräuschen, Anleitung zur Artikulation/prosodischen Gestaltung)* | Dlaska/Krekeler (2013) Hirschfeld/Reinke (2016) |
| explizit adaptiv | Möglichkeit der Wiederholung *(direkt (auf Empfehlung) oder indirekt durch Wiederholung der Übung (nachgelagert))* | Hirschi et al. (2022) |

[14] Unterstützungsangebote sind hinsichtlich metasprachlichen Wissens und kognitiver Anforderungen seitens der Nutzer:innen unterschiedlicher Qualität.

# Anhang 2

## Detailanalyse Ausspracheübung / Schematische Darstellung

Legende: < > = Inhalt variiert je nach Übung, [ ] = Inhaltsbeschreibung,
// = technische Einbettung, „“ = direktes Zitat

### (a) Online Aussprachetrainer

| Modus → | Schriftsprachlich | Visuell | Auditiv |
|---|---|---|---|
| ↓ Zeitlicher Ablauf | | | |
| (1) //Pop-Up | [Ziel der Übung und technische Erläuterung] | //Nachbau der Übung (clickable) | |
| (2) [Übung starten] | <Titel-Zielphänomen> | //Icon (Fragezeichen) = //Pop-Up – Einführende Erläuterungen | |
| | <Zielwort> | | //Artikulationsmodell (Audio) |
| | | //Icon (Fragezeichen) = //Pop-Up – Ziel der Übung | |
| (3) [Aufnehmen] | „Sprich nach!“ | //Icon (Mikrophon) in grün (pulsierend) | |
| (4.1) [Aussprache „korrekt“ ] | „Richtig!“ | <Zielphänomen> in grün //Icon (Haken) //Umrahmung in grün | |
| (4.2) [Aussprache „fast korrekt“] | „Das war fast richtig! Versuch es noch einmal!“ | <Zielphänomen> in rot //Umrahmung in gelb | |
| (4.3) [Aussprache „inkorrekt“ ] | „Versuch es noch einmal!“ | <Zielphänomen> in rot //Umrahmung in rot | |
| (5) [eigene Aufnahme anhören] | | //Icon (Lautsprecher) in grün | |
| (6) [nach zweitem Fehlversuch bei 4.2 oder 4.3] | <Phänomengeleiteter Artikulationstipp aus den einführenden Erläuterungen> | <Zielphänomen> in rot //Umrahmung in gelb oder rot | |
| (7) [Weiter] | //Button: „Überspringen“ | | |
| | //Button: „Weiter“ | | |
| (8) [Zusammenfassung] | „Geschafft“ | //Icon (Medaille, Kreuze, Haken) Farben grün, rot | |
| | [Wörter, Score, Anzahl Versuche] | | |

**(b) Sylby (segmental)**

| Modus → / ↓ Zeitlicher Ablauf | Schriftsprachlich | Visuell | Auditiv |
|---|---|---|---|
| (1) [Übung starten] | „Übung starten“ | //Button<br>in orange | |
| | <Zielwort / Zielsatz><br><IPA-Transkription> | | //Artikulationsmodell (Audio) |
| | | //Icon (Pfeil zurück)<br>= //Erklärvideo – Einführende Erläuterungen | |
| (2) [Aufnehmen] | „Mic-Taste antippen und halten zum Aufnehmen. Los geht’s!“ | //Icon (Mikrophon)<br>in orange | |
| (3.1) [Aussprache „korrekt“ ] | „Perfekt! Du hast ein Händchen für Sprache.“<br>in grün | <Zielphänomen(e)><br>in grün, fett, unterstrichen | |
| (3.2) [Aussprache „inkorrekt“ ] | „Noch nicht ganz gelungen. Versuch es doch nochmal.“<br>in rot | <Zielphänomen(e)><br>in rot, fett, unterstrichen | |
| (4) [Beispielaufnahme anhören] | „Beispiel“ | //Icon (Lautsprecher)<br>in orange | |
| (4) [eigene Aufnahme anhören] | „Du“ | //Icon (Lautsprecher)<br>in orange | |
| (4) [Aufnahme wiederholen] | | //Icon (Mikrophon)<br>in orange | |
| (4) [Navigation] | | //Icon (Pfeil rechts)<br>[weiter / überspringen] | |
| | | //Icon (Pfeil links)<br>[zurück] | |
| (5) [Zusammenfassung] | „Gut gemacht! Bleib dran und nimm die nächste Übung gleich in Angriff. <Score in Prozent> Gutes Ergebnis!“ | //Icon (Trophäe)<br>Farben orange, weiß | |
| | //Button: „Beenden“ | | |

**(c) Sylby (suprasegmental – „Intonation!? Intonation!“)**

| Modus → / ↓ Zeitlicher Ablauf | Schriftsprachlich | Visuell | Auditiv |
|---|---|---|---|
| (1) [Übung starten] | „Übung starten“ | //Button in orange | |
| | <Zielwort> in orange | //Grafik [Intonationsverlauf, abstrakte Darstellung des Modells in Form einer Geraden (gelb)]) | //Artikulationsmodell (Audio) |
| | | //Icon (Pfeil zurück) = //Erklärvideo – Einführende Erläuterungen | |
| (2) [Aufnehmen] | „Mic-Taste antippen und halten zum Aufnehmen. Los geht’s!“ | //Icon (Mikrophon) in orange | |
| (3) [Intonations-Ergebnis] | | //Grafik [eigener, nicht-abstrahierter Intonationsverlauf (grün) abgebildet parallel zur abstrakten Darstellung des Modells in Form einer Geraden (gelb)]) | |
| (4) [Beispielaufnahme erneut anhören] | „Beispiel“ | //Icon (Lautsprecher) in orange | |
| (4) [Navigation] | | //Icon (Pfeil rechts) [weiter / überspringen] | |
| | | //Icon (Pfeil links) [zurück] | |
| (5) [Zusammenfassung] | „Geschafft! Bleib dran und nimm die nächste Übung gleich in Angriff.“ | //Icon (Trophäe) Farben orange, weiß | |
| | //Button: „Beenden“ | | |