ASSISTUK

ASSISTUK GRUNDLAGEN · KI & TECHNIK

Solange die Stimme da ist

Message Banking, Stimmwahl und personalisierte Stimmen: was ein Werkzeug mit einer Stimme tun darf

Lars Tiedemann

Dipl. Heilpädagoge (FH) · AssistUK

17. November 2026

Arbeitsfassung. Der ALS-Text dieser Reihe sagt, die eigene Stimme lasse sich doppelt bewahren, aber nur, solange sie da ist. Dieser Text hier sagt, was die beiden Verfahren eigentlich sind, was sie nicht sind, und welche Regeln daraus für ein Werkzeug folgen. Er gilt nicht nur bei der ALS. Wer schon undeutlich spricht, wer nie eine lautsprachliche Biografie hatte, und wer eine Stimme nur auswählt, steht mit im Text. Alle Quellen wurden am 17.11.2026 auf Existenz und Kernaussage geprüft; das Verzeichnis weist den Prüfstatus aus. Wo ich eine Regel ableite, die selbst keine Studie ist, sage ich das. Der Schwestertext »Wessen Worte sind das?« behandelt die Urheberschaft von Sätzen. Dieser hier behandelt den Klang, in dem sie ankommen.

Inhalt

1. Drei Dinge, die ein Wort nicht unterscheiden kann
2. Die echten Sätze: Message Banking
3. Der Klang für neue Sätze: Voice Banking
4. Double Dipping, und warum es ein Angebot bleibt
5. Wenn die Stimme schon beeinträchtigt ist
6. Eine Stimme wählen, die nicht die eigene ist
7. Ton ist nicht dasselbe wie Stimme
8. Was ein Werkzeug mit einer Stimme tun darf
9. Grenzen dieses Textes
10. Fazit in fünf Sätzen
11. Literatur (mit Prüfvermerk)

---

1. Drei Dinge, die ein Wort nicht unterscheiden kann

Im Alltag heißt alles »die Stimme aufnehmen«. Dahinter stehen drei Verfahren, und sie beantworten drei verschiedene Fragen.

Message Banking sichert ganze Äußerungen: den Kosenamen in genau dieser Betonung, den Geburtstagsgruß, den Fluch, das Gebet. Abgespielt wird später die Aufnahme, nicht eine Nachbildung. Die Frage lautet: Bleibt dieser eine Satz sie selbst?

Voice Banking nimmt ein Korpus auf, damit daraus eine synthetische Stimme wird, die neue Sätze sprechen kann. Die Frage lautet: Klingen Sätze, die es beim Aufnehmen noch nicht gab, nach ihr?

Stimmwahl ist der dritte Fall, und er ist der häufigste. Die Person hatte nie die Lautsprache, die man hätte sichern können, oder sie will sie nicht sichern. Sie wählt unter vorhandenen Stimmen. Die Frage lautet dann nicht »ist es ihre alte Stimme?«, sondern »hat sie gewählt, und darf niemand sonst das für sie erledigen?«

Wer die drei vermischt, verspricht das Falsche. Eine Aufnahme ist keine Synthese. Eine Synthese ist kein Archiv. Und eine gewählte Stimme ist keine Fälschung, solange niemand behauptet, sie sei die verlorene.

2. Die echten Sätze: Message Banking

Das Vorgehen, das im Feld den Namen trägt, kommt aus dem Boston Children's Hospital. John Costello hat es als klinischen Weg beschrieben, Martine Smith und er haben es 2021 als Tutorial gefasst: Die Person nimmt über die Zeit Laute, Wörter, Sätze und Geschichten in der eigenen Stimme auf, mit der Betonung, die zu genau diesem Satz gehört. Es gibt keine Mindestzahl, weil nichts zerlegt und neu zusammengesetzt wird. Jede Aufnahme bleibt ein Ganzes und kann später an einem eigenen Platz im Gerät liegen (Costello & Smith 2021; das Verfahren selbst wird dort auf Costello 2000 zurückgeführt).

Der Wert ist nicht die Technik. Der Wert ist, dass ein Satz, der die Person ausmacht, nicht in eine freundliche Durchschnittsstimme übersetzt wird. Das Umfeld hört nicht »eine Stimme, die ungefähr passt«. Es hört sie.

Daraus folgt die Reihenfolge, die der ALS-Text schon setzt und die hier für jeden fortschreitenden Verlauf gilt: anbieten, solange die Stimme trägt. Nicht, wenn sie weg ist. Und anbieten, nicht verlangen. Die Aufnahme konfrontiert mit dem Verlauf. Wer sie zur Bedingung der Versorgung macht, hat das Verfahren gegen seinen eigenen Grundsatz gewendet. Costello und Smith stellen sechs Prinzipien unter das Vorgehen; der rote Faden, den sie ausdrücklich nennen, ist Eigentum und Steuerung bei der Person und ihrer Familie, nicht bei der Klinik und nicht beim Hersteller (Costello & Smith 2021).

3. Der Klang für neue Sätze: Voice Banking

Voice Banking beantwortet die andere Frage. Aus Aufnahmen wird eine phonetische Basis, aus der das Gerät Sätze sprechen kann, die nie aufgenommen wurden. Costello und Smith beschreiben den klassischen Weg so: ein ruhiger Raum, ein gutes Mikrofon, und je nach Dienst 1.600 oder mehr Phrasen. Neuere Dienste mit tiefen neuronalen Netzen kommen mit deutlich weniger aus; sie nennen Größenordnungen ab etwa 50 Aufnahmen. Beides hat denselben Haken. Auch eine gute synthetische Stimme trägt die Prosodie des Augenblicks nicht mit, die eine echte Aufnahme hat: den Ärger, die Zärtlichkeit, den Witz an genau dieser Stelle (Costello & Smith 2021).

Deshalb ist die Synthese kein Ersatz für die Bank, und die Bank kein Ersatz für die Synthese. Die Bank kann den Satz »Ich hab dich lieb« in ihrer Stimme sagen, und nur diesen. Die Synthese kann den Satz sagen, den es morgen erst gibt, und sie klingt nach ihr, ohne dieser eine Satz zu sein. Wer nur eines von beiden anbietet, hat die Hälfte der Aufgabe beschrieben.

4. Double Dipping, und warum es ein Angebot bleibt

Double Dipping ist der Brückenschlag: Die schon gebankten Botschaften werden zum Material der synthetischen Stimme, statt ein zweites, separates Skript aufzunehmen. Costello und Smith berichten aus der klinischen Erfahrung, dass Sprecherinnen und Sprecher diese

Stimmen der eigenen biologischen Stimme näher fanden als eine Stimme aus klassischem Voice Banking allein. Als Größenordnung für diesen Weg nennen Darstellungen des Verfahrens, die sich auf dieselben Autoren stützen, mehrere hundert Botschaften. Das ist Erfahrung aus der Praxis des Verfahrens, kein Wirksamkeitsnachweis mit Kontrollgruppe (Costello & Smith 2021).

Für ein Werkzeug heißt das: Double Dipping darf ein Schalter sein. Es darf keine Pflicht werden, »genug Sätze« zu schaffen, bevor die Person eine Stimme bekommt. Wer die Bank zur Rohstoffmine umdefiniert, verschiebt das Ziel. Die Aufnahme war ein Satz für später. Dass sie zusätzlich ein Modell füttern kann, ist ein Gewinn, den die Person freigibt oder nicht.

5. Wenn die Stimme schon beeinträchtigt ist

Die saubere Geschichte — erst aufnehmen, dann verlieren — gilt für einen Teil der Menschen und für einen anderen nicht. Bei angeborenen Bedingungen und bei einem plötzlichen Beginn gibt es kein unversehrtes Archiv. Und bei fortschreitenden Dysarthrien ist die Stimme oft schon verändert, wenn jemand zum ersten Mal über eine persönliche Stimme spricht.

Sarah Creer, Stuart Cunningham, Phil Green und Junichi Yamagishi haben 2013 gezeigt, dass man aus bereits beeinträchtigtem Sprechen eine personalisierte Synthese bauen kann: Eine Durchschnittsstimme, trainiert an unbeeinträchtigten Sprechern, wurde mit Aufnahmen von drei Menschen mit unterschiedlich schwerer Dysarthrie angepasst. In der qualitativen Einschätzung durch die Personen selbst und durch Hörer, die sie kannten, trug bei einer der drei die Stimme erkennbar Persönliches: regionale Färbung, Geschlecht, Alter. Bei den anderen beiden nicht in derselben Weise (Creer, Cunningham, Green & Yamagishi 2013). Drei Personen sind kein Versprechen. Sie sind die Grenze des Satzes »das geht«: Es geht bei manchen, und man erfährt es nur, indem man es versucht und die Person urteilen lässt.

Zwei Regeln folgen, und beide sind Ableitungen, keine Studienergebnisse. Erstens: Eine beeinträchtigte Stimme ist kein Ausschluss. Sie ist ein unsichereres Material, und das sagt man vorher. Zweitens: Die Person entscheidet, ob das Ergebnis sie ist. Nicht der Hörtest einer Fachkraft, und nicht die Ähnlichkeitszahl eines Dienstes.

6. Eine Stimme wählen, die nicht die eigene ist

Die meisten Menschen, für die wir Vokabulare und Zugänge bauen, werden keine Stimme aus der eigenen Lautsprache ziehen. Ein Kind, das nicht spricht, hat keine verlorene Erwachsenenstimme, die man restauriert. Ihm eine Stimme »passend zur Diagnose« zuzuweisen, ist dieselbe Bewegung wie ein Vokabular passend zur Diagnose: Die Person kommt in der Entscheidung nicht vor. Der Nachbartext »Die Diagnose ist kein Rezept« gilt auch für den Klang.

Stimmwahl ist deshalb eine Entscheidung der Person, in dem Tempo, das sie hat. Alter, Geschlecht, Akzent und Helligkeit sind keine klinischen Parameter. Sie sind Ausdruck. Wer für ein Kind wählt, weil »es schneller geht«, hat eine Entscheidung getroffen, die das Kind später als fremd erbt. Die Praxisableitung, an der ich unsere eigenen Werkzeuge messen will: Vorschlagen ja, festlegen nein. Eine Stimme lässt sich später wechseln, ohne dass Wörter, Plätze und Verlauf verloren gehen. Der Klang ist eine Einstellung, kein Besitzstand der Fachkraft.

Eine besondere Falle ist die Stimme der Mutter, des Vaters, der Assistenz als »seine Stimme«. Sie ist vertraut, und sie ist die falsche Person. Das Gerät spricht dann dauerhaft mit der Stimme von jemand anderem. Das ist keine Nähe. Das ist eine Besetzung.

7. Ton ist nicht dasselbe wie Stimme

Graham Pullin und Shannon Hennig haben 2015 beschrieben, was Sprachcomputer ihren Nutzerinnen und Nutzern fast vollständig vorenthalten: die Verfügungsgewalt über den Ton. Nicht die Stimmlage als Produktname, sondern die Art, wie ein Ja gesagt wird — bittend, scharf, belustigt, müde. In einer Übung mit vierzig Menschen aus dem Feld kamen, auf die Frage nach siebzehn Weisen, Ja zu sagen, über 250 verschiedene Beschreibungen zusammen. Sie sortierten sich nicht in eine Gefühlsskala, sondern in vier Blicke: Gefühlszustand, Absicht im Gespräch, soziales Verhältnis, stimmliche Qualität (Pullin & Hennig 2015).

Ihre drei Entwürfe — Stühle, eine Hecke aus Sprechweisen, ein Ton-Tisch — sind Kritik am Gerät, keine fertige Oberfläche. Ich ziehe daraus keine Bedienvorschrift. Ich ziehe eine Grenze: Eine persönliche Stimme, die jeden Satz im selben Ton sagt, hat die Identität zur Hälfte gelöst und den Ausdruck nicht. Wer einen Ton-Regler baut, muss ihn so bauen, dass er nicht langsamer ist als der Satz. Ein Ton, den man in einem Untermenü sucht, kommt zu spät. Das ist eine Ableitung aus dem, was Pullin und Hennig am Fehlen zeigen, und aus dem, was der Text über das Suchen auf Rastern zur Tiefe sagt. Es ist keine Evaluationsstudie.

8. Was ein Werkzeug mit einer Stimme tun darf

Die folgenden acht Regeln sind Praxisableitungen für die eigene Entwicklung. Sie sind keine Norm. Sie stehen hier, damit man sie an den Programmen prüfen kann.

1. **Getrennt halten.** Aufnahme, Synthese und Auswahl sind drei Schalter mit drei Sätzen darüber, was sie leisten und was nicht. Ein Knopf namens »Stimme sichern« ist keiner von den dreien.
2. **Die Aufnahme gehört der Person.** Sie liegt so, dass die Person sie mitnehmen kann, wenn sie den Dienst wechselt. Ein Hersteller, der die einzige Kopie hält, hält die Stimme.
3. **Löschen ist Teil des Verfahrens.** Was gebankt wurde, lässt sich einzeln und ganz entfernen. Auch Jahre später.

4. **Die Synthese spricht nur, was die Person gibt.** Stichwörter, eine Auswahl, eine Bestätigung — die Architektur steht im Text »Wessen Worte sind das?«. Eine Stimme, die von sich aus Sätze ergänzt, hat den Klang gelöst und die Urheberschaft aufgegeben. Der Ethik-Text zieht den Bruch nicht zwischen Mensch und Maschine, sondern dort, wo die Hilfe den Inhalt liefert und die Entscheidung mitnimmt.
5. **Double Dipping ist ein Angebot.** Es wird erklärt. Es wird nicht zur Eintrittsbedingung.
6. **Beeinträchtigtes Sprechen schließt nicht aus.** Es wird ehrlich als unsicher bezeichnet, und die Person urteilt über das Ergebnis.
7. **Zuweisen ist keine Hilfe.** Keine Stimme nach Diagnose, keine Stimme der Bezugsperson als Dauerlösung, kein Wechsel, der die Wörter verschiebt.
8. **Ton bleibt eine offene Aufgabe.** Bis es einen Weg gibt, der im Tempo der Äußerung liegt, behaupten wir nicht, Ausdruck sei mit der Stimmfarbe erledigt.

9. Grenzen dieses Textes

Message Banking ist ein publiziertes klinisches Vorgehen mit Prinzipien aus langer Praxis, keine kontrollierte Interventionsstudie. Die Aussage, double-dipping-Stimmen klängen näher, ist klinische Erfahrung der Autoren, kein Gruppenvergleich. Die Arbeit von Creer und Kollegen umfasst drei Personen und eine qualitative Einschätzung; sie belegt Möglichkeit, nicht Quote. Pullin und Hennig entwerfen, sie messen keine Nutzungsrate. Wie viele Aufnahmen ein bestimmter kommerzieller Dienst heute verlangt, ändert sich und gehört nicht in einen Grundlagentext. Die rechtliche Frage, wem eine synthetische Stimme nach dem Tod gehört, lasse ich offen; sie ist nicht durch die zitierten Arbeiten entschieden. Und ich baue selbst Werkzeuge, die Stimmen abspielen werden. Die acht Regeln sind zuerst die Messlatte für diese Werkzeuge.

10. Fazit in fünf Sätzen

Eine Stimme aufnehmen kann drei verschiedene Dinge meinen, und nur wer sie trennt, verspricht das Richtige: den echten Satz, den Klang für neue Sätze, oder eine Wahl unter Stimmen, die nie die eigenen waren. Die echten Sätze gibt es nur, solange die Stimme da ist, und sie bleiben Eigentum der Person — anbieten, nicht verlangen, mitnehmen können, löschen können. Die Synthese ergänzt die Bank, sie ersetzt sie nicht, und auch beeinträchtigtes Sprechen kann bei manchen Menschen noch Identität tragen, ohne dass daraus ein Versprechen wird. Wer nie lautsprachlich archiviert werden konnte, wählt selbst; eine Stimme nach Diagnose oder die Stimme der Mutter ist eine Besetzung. Und der Ton, in dem ein Ja gesagt wird, ist eine eigene Aufgabe, die eine schöne Stimmfarbe nicht erledigt.

11. Literatur (mit Prüfvermerk)

Alle Quellen am 17.11.2026 auf Existenz und Kernaussage geprüft (✓). Praxisregeln in Abschnitt 8 sind als Ableitung gekennzeichnet und keine zusätzlichen Quellen.

1. ✓ **Costello, J. M. & Smith, M. (2021).** The BCH message banking process, voice banking, and double-dipping. *Augmentative and Alternative Communication*, 37(4), 241–250. DOI 10.1080/07434618.2021.2021554. — Tutorial zum Boston-Vorgehen: ganze Äußerungen in der eigenen Prosodie; Voice Banking als phonetische Basis für neue Sätze; Double Dipping als optionale Brücke; Eigentum und Steuerung bei der Person. Das Verfahren selbst führen die Autoren auf Costello (2000) zurück.
2. ✓ **Creer, S., Cunningham, S., Green, P. & Yamagishi, J. (2013).** Building personalised synthetic voices for individuals with severe speech impairment. *Computer Speech & Language*, 27(6), 1178–1193. DOI 10.1016/j.csl.2012.10.001. — Anpassung einer Durchschnittsstimme mit dysarthrischen Aufnahmen von drei Personen; bei einer blieben regionale Färbung, Geschlecht und Alter für vertraute Hörer erkennbar.
3. ✓ **Pullin, G. & Hennig, S. (2015).** 17 ways to say yes: Toward nuanced tone of voice in AAC and speech technology. *Augmentative and Alternative Communication*, 31(2), 170–180. DOI 10.3109/07434618.2015.1037930. — Ton als fehlende Verfügungsgewalt; über 250 Beschreibungen aus einer Übung mit 40 Personen, vier Perspektiven statt einer Gefühlsskala. Entwürfe, keine fertige Oberfläche.
4. ✓ **Light, J. (1989).** Toward a definition of communicative competence for individuals using augmentative and alternative communication systems. *Augmentative and Alternative Communication*, 5(2), 137–144. DOI 10.1080/07434618912331275126. — Kommunikative Kompetenz als Zusammenspiel von sprachlicher, operativer, sozialer und strategischer Kompetenz. Hier nur für die strategische Seite: eine Stimme nützt nichts, wenn der Weg zu ihr die Äußerung verschluckt.
5. ✓ **Tiedemann, L. (2026).** Wenn die Sprache geht (ALS) · Wessen Worte sind das? (Urheberschaft) · Hilfe, die nicht übernimmt (Ethik der Hilfen) · Hicks Gesetz gilt hier nicht (Suchen und Tiefe). AssistUK-Grundlagenreihe. — Die Nachbartexte: der Zeitpunkt, die Sätze, die Leiter der Eingriffstiefe, die Kosten einer zusätzlichen Ebene.

Die fachliche Position, Auswahl und Verantwortung liegen bei mir. — L. T.