



Bern, 26. März 2025

# Die automatische Transkription von Vernehmungen mittels KI

Schweizer Polizei Informatik Kongress 2025

Dr. ès. sc. David Imseng, recapp IT AG, Visp

Dr. iur. Andreas Kind, Kantonale Staatsanwaltschaft Wirtschaftsdelikte, Bern

# Geschwindigkeit: Sprache vs. Tippen



130 Wörter / Min.

≅ rund 800 Zeichen / Min.



Fortgeschrittene: 200 - 400 Anschläge / Min.



# Die revidierte Strafprozessordnung (seit 2024 in Kraft)

## Art. 78a<sup>32</sup> Einvernahmeprotokolle bei Aufzeichnung der Einvernahme

Wird die Einvernahme mit technischen Hilfsmitteln aufgezeichnet, so gelten gegenüber den allgemeinen Regeln (Art. 78) folgende Abweichungen:

- a. Anstelle einer laufenden Protokollierung während der Einvernahme kann das **Protokoll auch erst danach gestützt auf die Aufzeichnung erstellt werden**, grundsätzlich jedoch innerhalb von sieben Tagen nach der Einvernahme.
- b. Die einvernehmende Behörde **kann darauf verzichten**, der einvernommenen Person das Protokoll vorzulesen oder zum Lesen vorzulegen und von dieser **unterzeichnen und visieren zu lassen**.
- c. Die Aufzeichnung der Einvernahme wird sofort zu den Akten genommen.

Die revidierte Strafprozessordnung (seit 2024 in Kraft)

Pro Minute Audio: 2-4 Minuten Tippen

Laborbedingungen: Faktor 1:2 bis 1:4

Realität: Faktor 1:8 bis 1:12

Wird die Einvernahme über den allgemeinen Regeln (Art. 78) folgende Abweichungen:

- a. Anstelle einer laufenden Protokollierung während der Einvernahme kann das **Protokoll auch erst danach gestützt auf die Aufzeichnung erstellt werden,** grundsätzlich jedoch innerhalb von sieben Tagen nach der Einvernahme.
- b. Die einvernehmende Behörde **kann darauf verzichten,** der einvernommenen Person das Protokoll vorzulesen oder zum Lesen vorzulegen und von dieser **unterzeichnen und visieren zu lassen.**
- c. Die Aufzeichnung der Einvernahme wird sofort zu den Akten genommen.

# Prozess bei der STAWA WD Bern

## Einvernahme Ablauf

- Aufnahme starten
- Rahmenprotokoll
  - Rubrum
  - Belehrung
  - Besondere Vorkommnisse
- «*Zur Sache: siehe Transkript*»



Einvernahme XY\_Datum.mp3



# Einvernahmearten im Vergleich



Kriterium	Audiovisuelle EV	Klassische EV
Dauer	✓	✗ <i>Unterbrüche, vorlesen, rückübersetzen</i>
Befrager:in	✓	✗ <i>kognitive Überforderung</i>
Qualität Protokoll	✓	✗ <i>weglassen, zusammenfassen</i>
Gutachten/Schulung	✓	✗
Protektiver Effekt	✓	✗
Nachbearbeitung	✗ <i>aufwendig</i>	✓



# Arbeitsweise mit voscriba

## Nachbearbeitung effizienter gestalten



# Upload

<https://youtu.be/GRp04AbBXBU>





# Arbeiten mit voscriba

<https://youtu.be/xjj2nKSrL1E>



# Word-Export

<https://youtu.be/sy6s5t6H19A>



# Learnings aus dem Pilotprojekt

# Mehrwert Audiovisuelle EV

## Reduktion Einvernahmedauer

Die Einvernahmedauer ohne parallel laufende Protokollierung kann deutlich um mehr als die Hälfte gekürzt werden.

⇒ erhöht Bereitschaft zu detailliertem freien Bericht

**Erfahrungsbsp.** (gleicher Beschuldigter, gleiches Thema, gleiche EV-«Crew»):

«klassisch»: 110 Min. / 4 Vorhalte ⇒ 27.5 Min./Vorhalt

mit Aufnahme: 97 Min. / 29 Vorhalte ⇒ 3.5 Min./Vorhalt

# Gesamtbetrachtung

Manuell: 1:8 bis 1:12 (Verhältnis Aufnahmedauer / Nachbearbeitungsaufwand)

KI: 1:3 bis 1:6

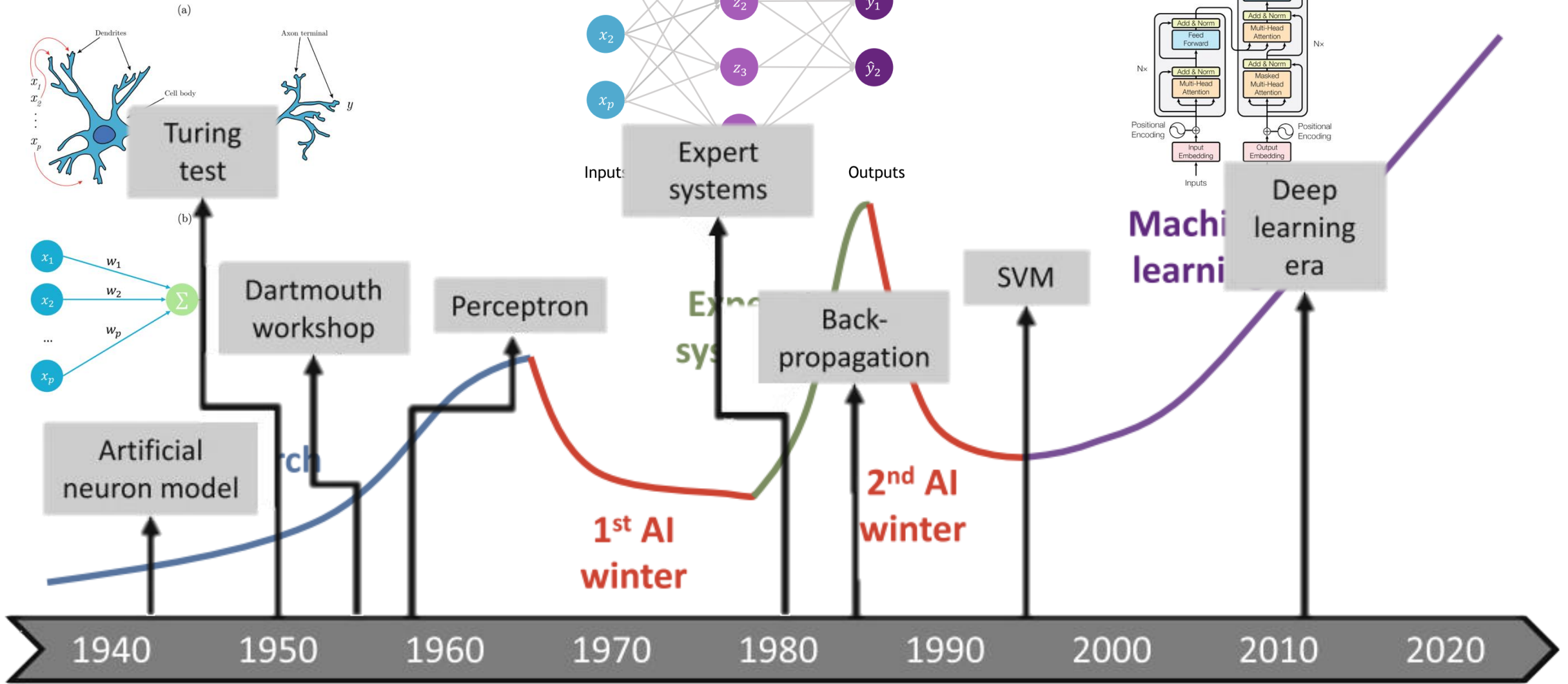
**Reduktion Nachbearbeitungszeit  
plus  
sämtliche  Vorteile einer audiovisuellen EV**

⇒ **Spracherkennung als Eisbrecher für audiovisuelle Einvernahmen?**



# Künstliche Intelligenz

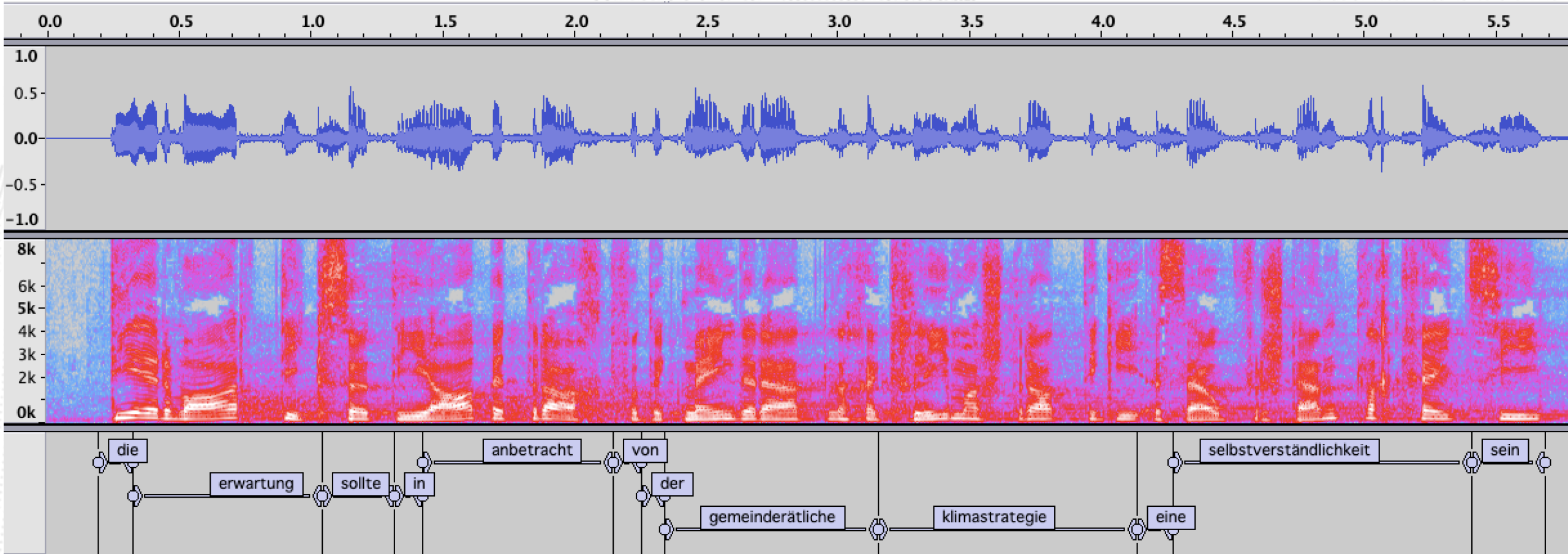
## Beispiel Spracherkennung



Aus: Colliot, O. (2023). A Non-technical Introduction to Machine Learning. In: Colliot, O. (eds) Machine Learning for Brain Disorders. Neuromethods, vol 197. Humana, New York, NY. [https://doi.org/10.1007/978-1-0716-3195-9\\_1](https://doi.org/10.1007/978-1-0716-3195-9_1)

Vaswani A. (2017). Attention is all you need. In: Processings of Advances in neural information processing systems, 5998-6008

# Das Spektrogramm

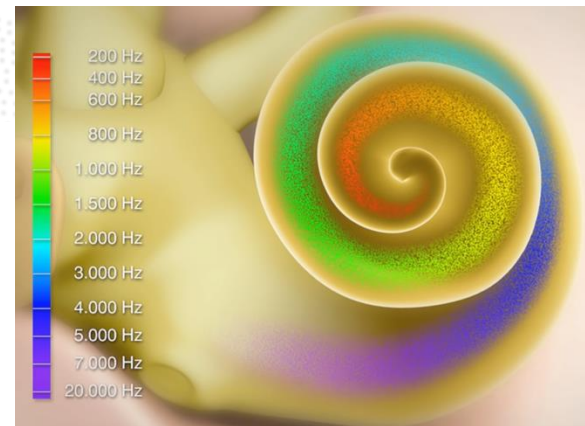




# Das Spektrogramm

Das Spektrogramm ist eine bildliche Darstellung des Frequenzspektrums einer Audioaufnahme  
grau/blau = wenig Energie  
rot/weiße = viel Energie

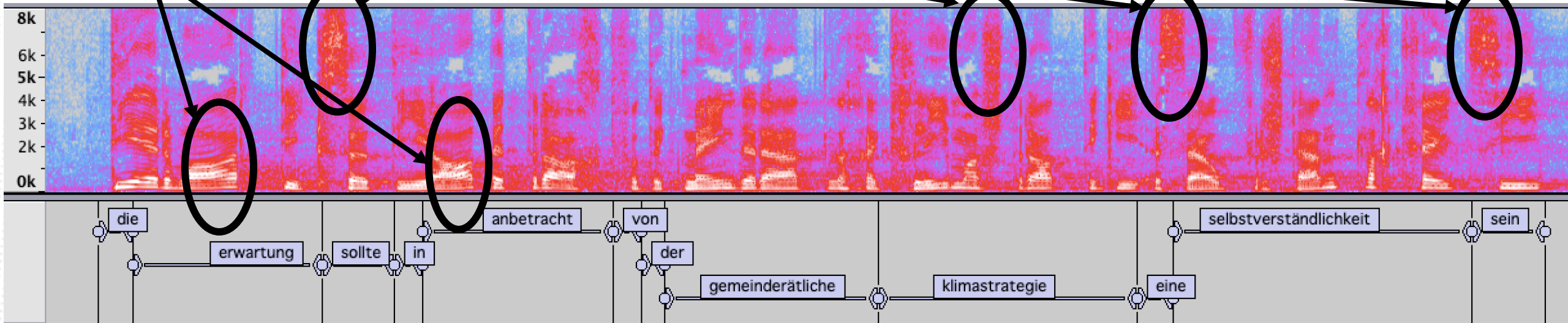
Der Mensch hört verschiedene Tonhöhen (=Frequenzen).



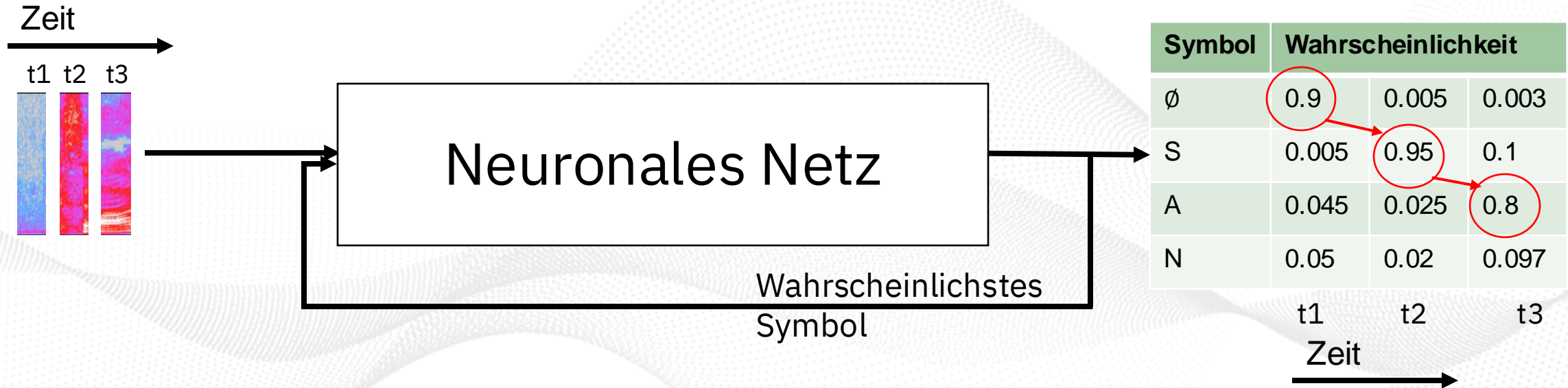
<https://www.medel.com/de-at/about-hearing/how-hearing-works>

”a”

”s”



# Der wahrscheinlichste Laut



Früher wurde jeder Audioabschnitt einzeln betrachtet.

Das ist in der Realität nicht so,

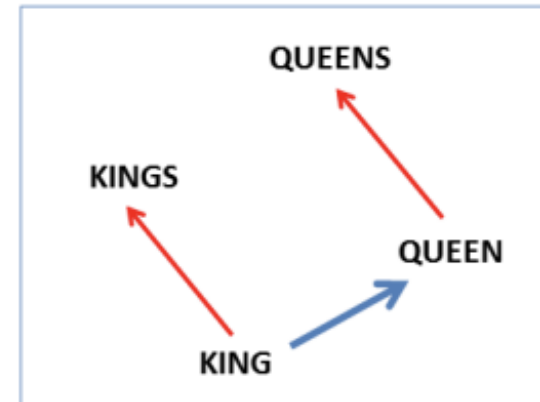
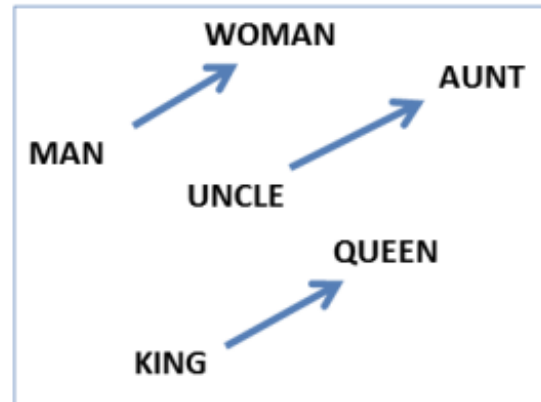
war aber aufgrund der Rechenkapazität eine notwendige Vereinfachung.

Heute werden die bereits erkannten Symbole (Laute) vom System berücksichtigt.

# Spracherkennung braucht viel Kontext

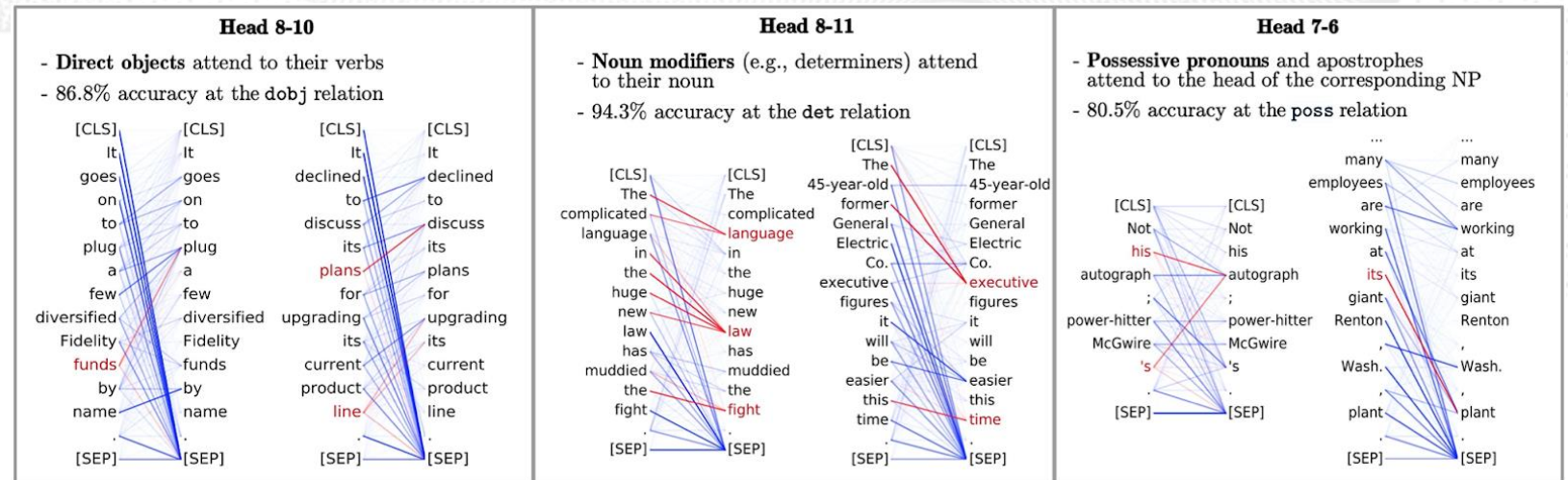


## Embeddings



Linguistic Regularities in Continuous Space Word Representations – Mikolov et al. 2013

## Attention is all you need



What Does BERT Look At? An Analysis of BERT's Attention - Clark et al. 2019



# Welche Herausforderungen bleiben?

- Sehr leise / sehr emotionale Äusserungen
- Junge Sprecher / Kinder
- Schlechte Formulierungen
- Schnelle Sprecherwechsel / durcheinanderreden
- Schlechte Audioqualität / Distanz zum Mikrofon

# Angepasste Open-Source Modelle

Modell	Fehlerquote (WER) auf realen Daten mit Jugendlichen
Whisper large (v2)	54.08 %
Whisper large (v3)	51.74 %
Angepasstes Modell	34.42 %


Ca. 35% relative Verbesserung  
(nur noch 65% der Fehler)

Videos zum mithören: <https://youtu.be/CMj5UbFUgGs>

## Whisper large v3

00:07:21  spkr-004 SRF Kids Reporterin. Du und deine Story. [...] Äh, müssen wir hier durch? Ja, okay, [...] gut.

## Angepasstes Modell

00:07:21  spkr-004 SRF Kids Reporter Inn. Du und deine Story. [...] Ähm, müssen wir hier durch? Ja, okay, gut. Luke, komm! Er hat keinen Platz mehr hinter uns, nicht genug, machen wir den Stau. Mister, kommen Sie bitte! Hallo! Luke, komm!



# Welche Herausforderungen bleiben?

- Sehr leise / sehr emotionale Äusserungen -> *Anpassungen*
- Junge Sprecher / Kinder -> *Anpassungen*
- Schlechte Formulierungen -> *Wie soll das Protokoll aussehen?*
- Schnelle Sprecherwechsel / durcheinanderreden -> *Forschungsthema*
- Schlechte Audioqualität / Distanz zum Mikrofon -> *Hardware / Mikrofone*

# Danke für die Aufmerksamkeit

## Fragen

Andreas Kind: [andreas.kind@justice.be.ch](mailto:andreas.kind@justice.be.ch)

David Imseng: [david.imseng@recapp.ch](mailto:david.imseng@recapp.ch)