

Bensheim Forum
14. Juni 2023

DNN-Training

Wie ein Hörsystem hören lernt

Wieviele Menschen braucht das Digitale?



DNN-Training

Wie ein Hörsystem hören lernt

- Motivation
- Beschreibung Chat GPD
- Training in Eriksholm
 - Grundlagenforschung
 - Technische Entwicklung
 - Training
 - Finalisierung
 - Überraschungen
- Ergebnisse



DNN-Training

Wie ein Hörsystem hören lernt

Motivation:

- DNN = Deep Neural Network – Tiefes neuronales Netzwerk
- Künstliche Intelligenz höchster Güte
- Kann lernen, wie das menschliche Gehirn
- Nach Training erstaunliche Fähigkeiten:
 - Autonomes Fahren 1968
 - Autonomes Fahren 2021



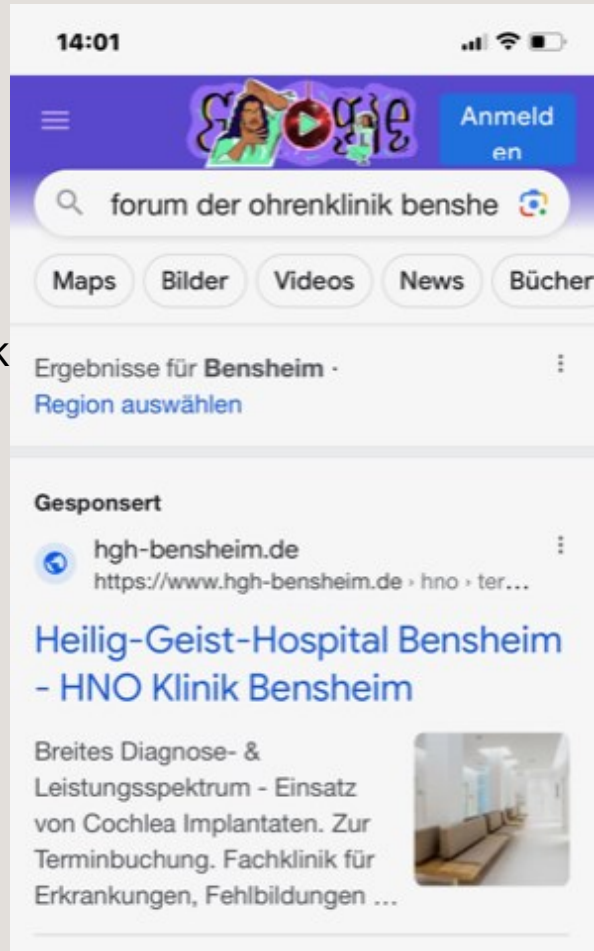
Foto: Hamburger Hochbahn

DNN-Training

Wie ein Hörsystem hören lernt

Motivation:

- DNN = Deep Neural Network – Tiefes neuronales Netzwerk
- Künstliche Intelligenz höchster Güte
- Kann lernen, wie das menschliche Gehirn
- Nach Training erstaunliche Fähigkeiten:
 - Autonomes Fahren 1968
 - Autonomes Fahren 2021
 - Suchmaschinen



DNN-Training

Wie ein Hörsystem hören lernt

Motivation:

- DNN = Deep Neural Network – Tiefes neuronales Netzwerk
- Künstliche Intelligenz höchster Güte
- Kann lernen, wie das menschliche Gehirn
- Nach Training erstaunliche Fähigkeiten:
 - Autonomes Fahren 1968
 - Autonomes Fahren 2021
 - Suchmaschinen
 - Wettervorhersage



DNN-Training

Wie ein Hörsystem hören lernt

Motivation:

- DNN = Deep Neural Network – Tiefes neuronales Netzwerk
- Künstliche Intelligenz höchster Güte
- Kann lernen, wie das menschliche Gehirn
- Nach Training erstaunliche Fähigkeiten:
 - Autonomes Fahren 1968
 - Autonomes Fahren 2021
 - Suchmaschinen
 - Wettervorhersage
 - Bildgebende Verfahren in der Medizin



<https://www.spektrum.de/wissen/was-koennen-hirnschans-und-was-nicht/1491877>

DNN-Training

Wie ein Hörsystem hören lernt

Motivation:

- DNN = Deep Neural Network – Tiefes neuronales Netzwerk
- Künstliche Intelligenz höchster Güte
- Kann lernen, wie das menschliche Gehirn
- Nach Training erstaunliche Fähigkeiten:
 - Autonomes Fahren 1968
 - Autonomes Fahren 2021
 - Suchmaschinen
 - Wettervorhersage
 - Bildgebende Verfahren in der Medizin
- Sprache und Umgebungsgeräusche wird ausbalanciert wiedergeben



DNN-Training

Wie ein Hörsystem hören lernt

Motivation:

- DNN ist super schnell und präzise, wenn es gut trainiert ist.

Problem:

- Das Training dauert lange.
- Es kann auch falsch oder nicht ausreichend trainiert sein.

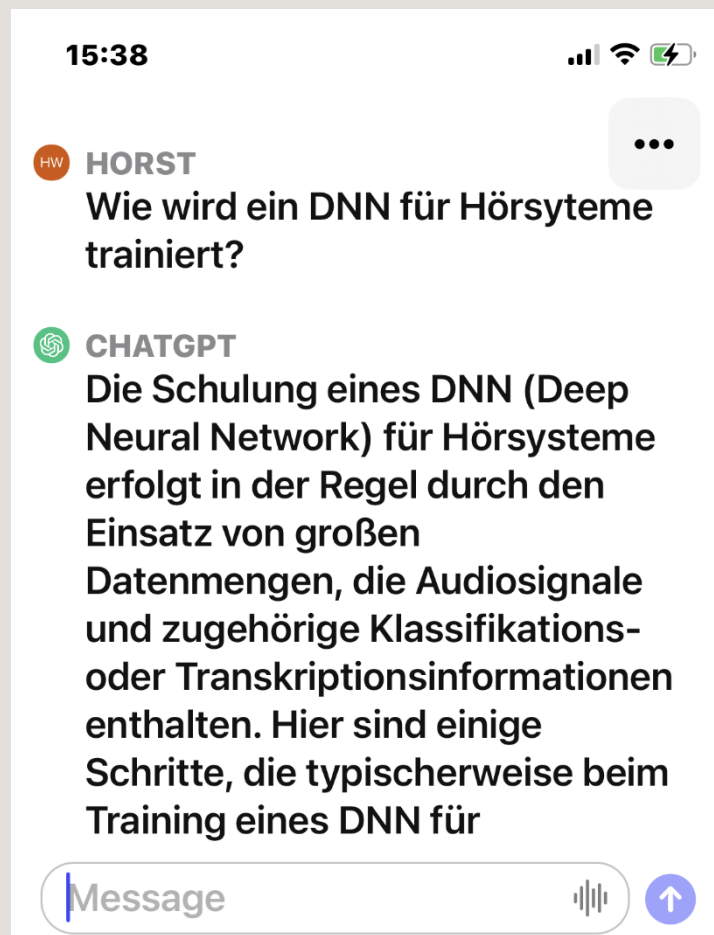


<https://www.faz.net/aktuell/wirtschaft/digitec/tesla-prallt-mit-autopilot-in-polizeiwagen-15614862.html>

DNN-Training

Wie ein Hörsystem hören lernt

Beschreibung Chat GPD:



DNN-Training

Wie ein Hörsystem hören lernt – Chat GPD:

1. Datensammlung: **Es werden umfangreiche Audioaufnahmen von verschiedenen Geräuschen und Sprachsignalen gesammelt.**
Diese Aufnahmen können sowohl in kontrollierten Umgebungen als auch in realen Situationen erfolgen.
2. Datenverarbeitung: **Die gesammelten Audiodaten werden in ein geeignetes Format umgewandelt, das für das Training des DNN geeignet ist.**
Dies kann beispielsweise eine spektrale Darstellung des Audiosignals sein, wie beispielsweise die Berechnung von Mel-Frequency Cepstral Coefficients (MFCCs).
3. Architekturdesign: **Es wird eine geeignete DNN-Architektur für das Hörsystem entwickelt.**
Dies kann ein Feedforward-DNN oder ein rekurrentes, neuronales Netzwerk (RNN) wie ein Long Short-Term Memory (LSTM) sein.
4. Training: **Das DNN wird mit den vorverarbeiteten Audiodaten trainiert.**
Dies geschieht durch die Eingabe der Audiodaten in das Netzwerk und die Anpassung der Gewichte und Parameter des Netzwerks, um die gewünschten Ausgabeergebnisse zu erzielen. Das Training erfolgt normalerweise mithilfe von Algorithmen wie dem Backpropagation-Verfahren und Gradientenabstiegsverfahren.

DNN-Training

Wie ein Hörsystem hören lernt – Chat GPD

5. Validierung und Optimierung: **Das trainierte DNN wird auf einem separaten Validierungssatz von Audiodaten getestet, um seine Leistung zu bewerten.** Basierend auf den Validierungsergebnissen können Anpassungen am DNN und am Trainingsprozess vorgenommen werden, um die Leistung zu optimieren
6. Evaluierung: **Das trainierte DNN wird auf neuen, unbekannten Audiodaten getestet, um seine Leistung im echten Einsatz zu bewerten.** Dies ermöglicht es, die Wirksamkeit des Hörsystems zu bewerten und weitere Verbesserungen vorzunehmen, falls erforderlich.

Es ist wichtig anzumerken, dass dies **eine vereinfachte Beschreibung des Trainingsprozesses** ist und dass die tatsächliche Implementierung und das Training eines DNN für Hörsysteme je nach den spezifischen Anforderungen und Zielen variieren können.

DNN-Training

Wie ein Hörsystem hören lernt – Chat GPD

5. Validierung und Optimierung: **Das trainierte DNN wird auf einem Validierungssatz von Audiodaten getestet, um seine Leistung zu bewerten.** Während auf den Validierungsergebnissen können Anpassungen vorgenommen werden, um die Leistung zu optimieren.
6. Evaluierung: **Das trainierte DNN wird auf einem Testset von Audiodaten getestet, um seine Leistung zu bewerten.** Möglichst es, die Wirksamkeit des Hörsystems zu bewerten, falls erforderlich.

Es ist wichtig, eine detaillierte Beschreibung des Trainingsprozesses zu haben, und das Training eines DNN für Hörsysteme je nach den spezifischen Anforderungen variieren können.

Wo ist der Mensch?

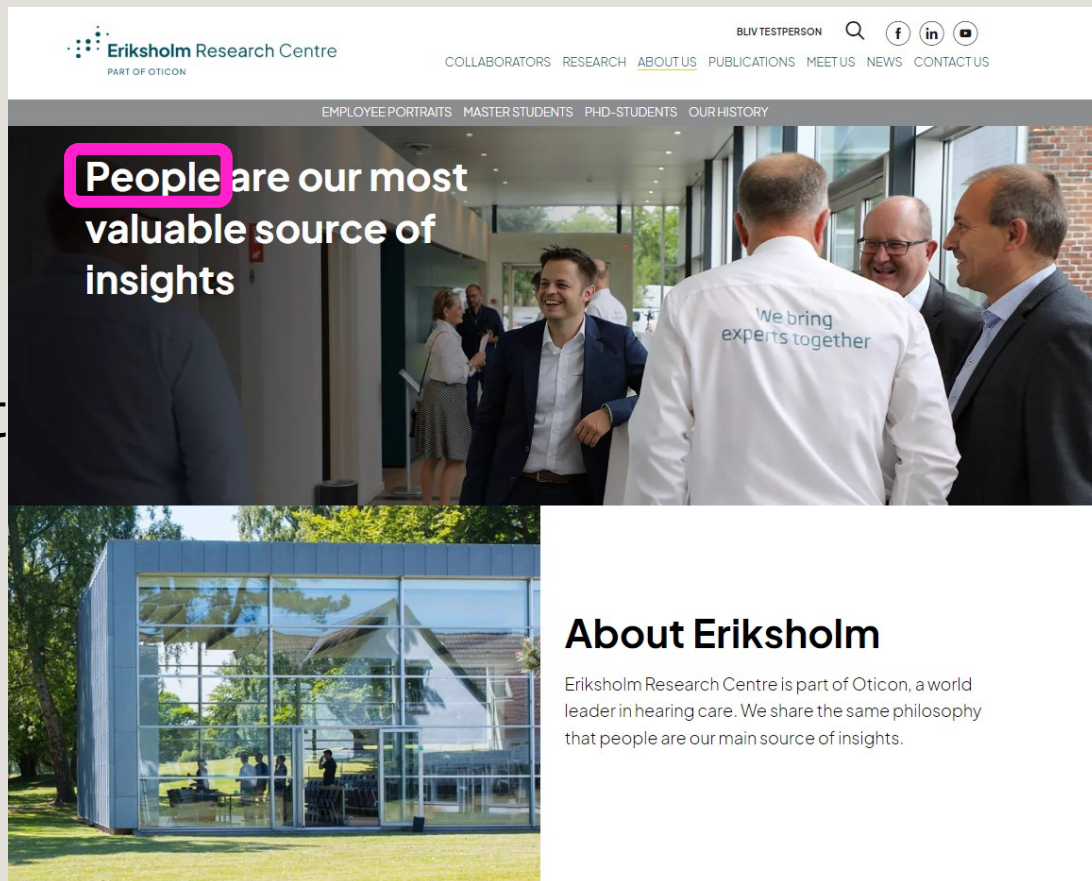
DNN-Training

Wie ein Hörsystem hören lernt





Training in Eriksholm

KI seit 2004 im Markt

DNN ab 2012 entwickelt



Eriksholm Research Centre
PART OF OTICON

BLIV TESTPERSON    

COLLABORATORS RESEARCH ABOUT US PUBLICATIONS MEET US NEWS CONTACT US

EMPLOYEE PORTRAITS MASTER STUDENTS PHD-STUDENTS OUR HISTORY

People are our most valuable source of insights

We bring experts together

About Eriksholm

Eriksholm Research Centre is part of Oticon, a world leader in hearing care. We share the same philosophy that people are our main source of insights.

DNN-Training

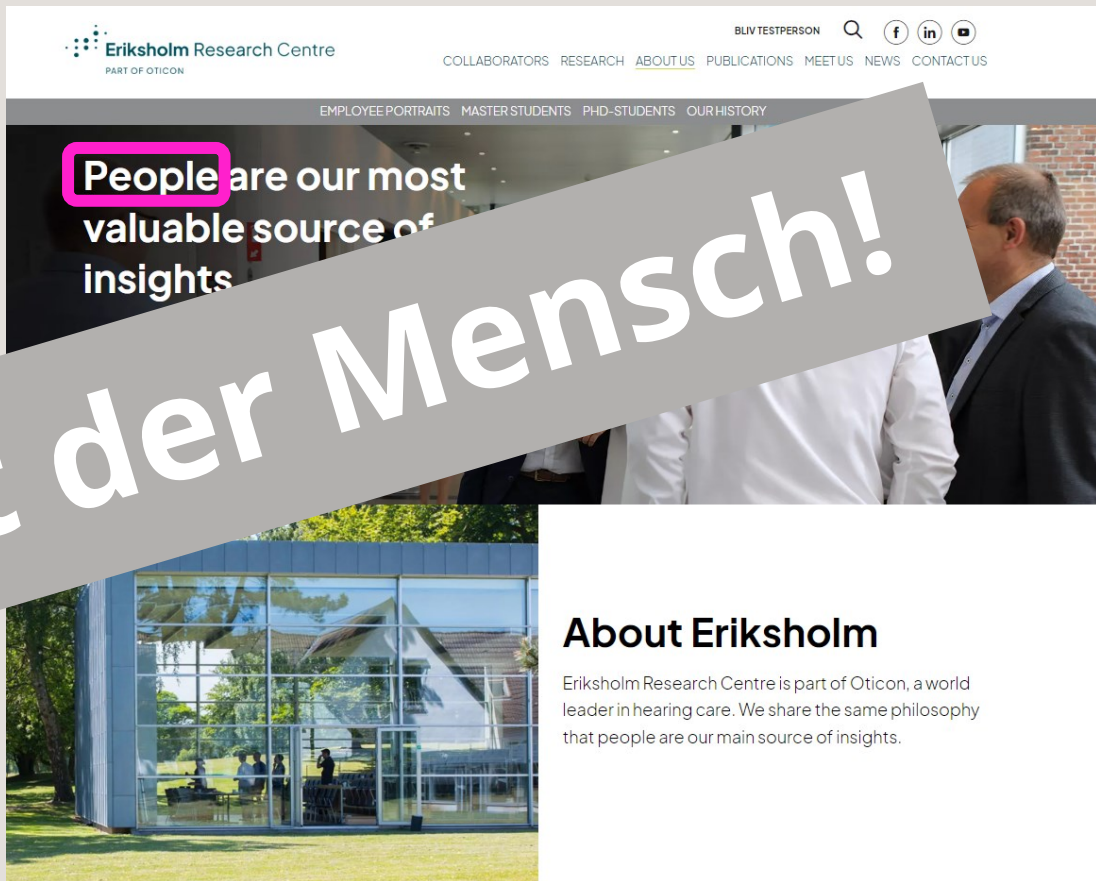
Wie ein Hörsystem hören lernt

Training in Eriksholm

KI seit 2004 im Markt

DNN ab 2012 entwickelt

Hier ist der Mensch!





DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

[< BACK](#)

Studies within Speech segregation using deep neural networks

 Does DNN separation of voices help hearing aid users segregate speakers?



Lars Bramsløw

Research Engineer, PhD, Project Leader

[> Learn more](#)

mail@eriksholm.com

Speech segregation using Deep Neural Networks (DNN)

Many hearing aid users struggle to segregate voices and this complicates social situations. Researchers from Eriksholm Research Centre aim at solving this by training hearing aids by using Deep Neural Networks (DNN).

What is DNN?

Deep Neural Networks (DNN) is a trainable algorithm that enables a computer to solve many human-like tasks like separating sounds, faces, and voices. As an example, the technology is already used in Smartphones to identify people in photographs and in automatic speech recognition. DNN works like large networks of simple 'neurons' inspired by the way the brain works. Each neuron is a simple element that receives multiple inputs and performs a series of simple actions which generate an output. When many neurons are stacked and densely interconnected, they can handle more complex input, and such neuron-layers can learn to perform very difficult tasks. This means that they learn from examples, and apply their knowledge to different, but similar, tasks. As an example, they can learn to analyse, recognise and separate specific sound sources.

DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Niels Henrik Pontoppidan
Group Manager
Eriksholm Research Centre

You don't teach it exactly how to do it
but you teach it what to do

DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Low-Latency Sound Source Separation Using Deep Neural Networks

Gaurav Naithani¹, Giambattista Parascandolo¹, Tom Barker¹, Niels Henrik Pontoppidan², and Tuomas Virtanen¹



¹Tampere University of Technology, Finland, ²Eriksholm Research Centre, Oticon A/S, Denmark



Overview

The monoaural source separation method is applied to two talker mixtures using feedforward deep neural networks (DNN) with no prior information other than the identity of speakers. The proposed approach is focused at low algorithmic delay applications, e.g., hearing aids. Around 1-2 dB improvement in source to distortion ratio (SDR) compared to non negative matrix factorization baseline are achieved.

- Low- algorithmic delay is paramount for real time applications. For example, in hearing aids even delays ≈ 20 ms results in listener discomfort.
- DNNs models the source separation task as non linear regression between input (mixture spectrum) and output (constituent source spectra or intermediate time-frequency masks).
- DNNs are better equipped to handle this task in comparison to compositional model based approaches, e.g., non negative matrix factorization (NMF).

Method

- Spectral short-time Fourier transform (STFT) features derived from two talker acoustic mixtures are used as DNN input to estimate time-frequency masks corresponding to individual speakers.
- Algorithmic latency as low as 5 ms have been achieved.

Time-frequency masking

- Soft time-frequency masks are used:

$$M(t, f) = \frac{|S_1(t, f)|}{|S_1(t, f)| + |S_2(t, f)|}$$

where S_1 and S_2 are spectral features of corresponding constituent sources.

Source reconstruction

- Individual source spectra are calculated from estimated DNN output M_{est} , as,

$$S_{est1} = M_{est}(t, f) * Y(t, f)$$

and

$$S_{est2} = (1 - M_{est}(t, f)) * Y(t, f)$$

where $Y(t, f)$ is the mixture spectrum.

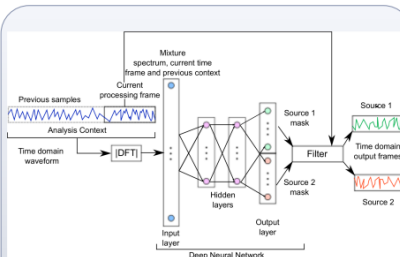


Fig.1. Proposed DNN based sound source separation approach.

- Features derived from a larger past temporal context used to predict time-frequency masks for current frame.
- Three layer feedforward DNN with 256 neurons in each layer is used.
- Neural network hyper parameters are chosen based on a validation set different from training and test sets.

2016

Acoustic Material

- CMU Arctic dataset A for training/validation and B for testing.
- Five speaker pairs: two male-male, two male-female and one female-female speaker pairs.
- 1024 acoustic mixtures for training, and 100 acoustic mixtures for testing for each speaker pair.

Metrics

- Source to distortion ratio (SDR), Source to Interference ratio (SIR), Source to artifact ratio (SAR)

Results

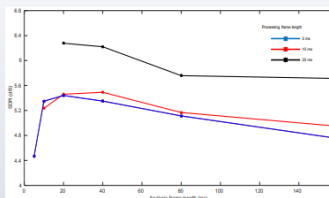


Fig.2. Variation of separation performance with analysis frame lengths for different processing frame lengths.

- Significant improvement in separation performance at low processing frame lengths.
- For larger processing frame lengths, using previous context is not of much help.
- Consistent improvement in separation performance over NMF baseline over all processing frame lengths.
for 5ms, SDR improvement over NMF is 1.8 dB.
for 10 ms, SDR improvement over NMF is 1.5 dB.

Conclusion

- A DNN based single channel source separation method for two talker mixtures has been proposed for low algorithmic delay applications.
- The effect of duration of the incorporated past temporal context on separation performance has been studied.
- The DNN based approach consistently outperforms NMF baseline for all latencies.
- Improvement in separation performance is most significant for very short processing frame lengths.

Contact details

gaurav.naithani@tut.fi

DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Training the DNN with the truth

3 min of clean speech

The diagram illustrates the training process for a Deep Neural Network (DNN) to reconstruct speech. It starts with a 'Deep Neural Network' block. Two paths emerge from it, each passing through a 'Time frequency masking' block (represented by a blue box and a microphone icon). The first path leads to a 'source1 spectrogram' (blue box), which then undergoes 'time domain reconstruction' to produce 'source1' (red waveform). The second path leads to a 'source2 spectrogram' (blue box), which undergoes 'time domain reconstruction' to produce 'source2' (black waveform). A large blue arrow points from the spectrograms back to the DNN, indicating a feedback loop. Above the spectrograms, a small diagram shows the DNN's internal structure: 'Input' enters a 'DNN' block, which produces an 'Output' that is compared against a 'Target' in a 'compare' block. The Eriksholm Research Centre logo is visible at the bottom right of the slide.

Deep Neural Network

Time frequency masking

source1 spectrogram

time domain reconstruction

source1

Time frequency masking

source2 spectrogram

time domain reconstruction

source2

Eriksholm Research Centre

DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Training mit reinen Sounds:

- Sprechern
- Geräuschen
- Konkurrierenden Sprechern

Wichtig: Realistische Sounds

- Sphärische Mikrofone
- Soundstudios



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Trainer, Tester - **Menschen:**

- Expert Listener
 - Normalhörend für “guten Sound”
- Expert Users
 - “As many as possible”
 - Mit Hörverlust für erlebten Nutzen
 - Gruppengröße jeweils ~ 30
 - Testzeiten jeweils ~ 2 Wochen



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Trainer, Tester - **Menschen:**

- Expert Listener
 - Normalhörend für “guten Sound”
- Expert Users
 - “As many as possible”
 - Mit Hörverlust für erlebten Nutzen
 - Gruppengröße jeweils ~ 30
 - Testzeiten jeweils ~ 2 Wochen

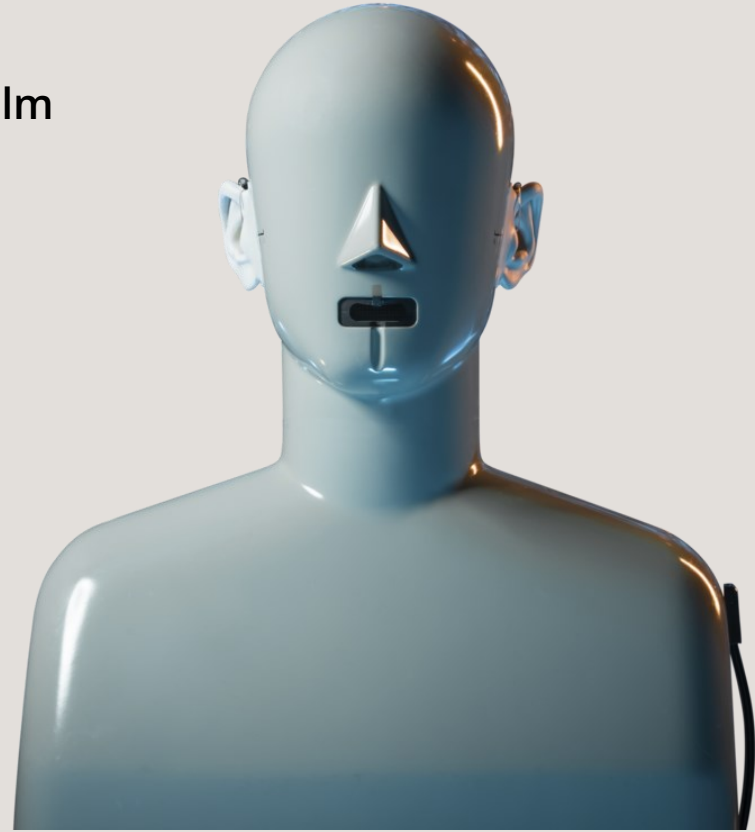


DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Trainer, Tester:

- “Robot Users”
 - HATS: **H**at **A**nd **T**orso **S**imulator
 - Sicherheit des Systems testen
 - Mit großen Datenmengen
 - Klangwelten, Sprache in Sprache



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Hardware:

- Notebook mit Kabel zu Hörsystemen
 - Hunderte DNN-Versionen!
- “Prototype Instrument”
 - ~ 30 DNN-Versionen



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Hardware:

- Notebook mit Kabel zu Hörsystemen
 - Hunderte DNN-Versionen!
- “Prototype Instrument”
 - ~ 30 DNN-Versionen
- “Real Instrument”
 - 2019
 - 2020 im Markt
 - 2023 nächste Generation



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Überraschungen:

- Der **Mensch** hört kleinste Unterschiede
 - Bei technisch “gleichen” Funktionalitäten
 - Unterschiedliche DNN-Versionen
 - Technische Messungen wertlos???



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Ergebnisse:

- Natürlicher Klang
- Schnelle Reaktion (1ms)
- Sofort erlebter Nutzen – vom **Menschen**



DNN-Training

Wie ein Hörsystem hören lernt – Training in Eriksholm

Ergebnisse, über 20 Studien, u.A.:

- 1) Bis zu 10 x schneller als Vergleichsgeräte
- 2) 60 % bessere Orientierung
- 3) Objektive Senkung Höranstrengung um 30 %
- 4) Subjektive Klangbewertung zu 80 % besser
- 5) 20 dB bessere Windgeräusch-Unterdrückung, als Vorgängermodell
- 6) SNR für Sprache um 4,3 dB verbessert, gegenüber Vorgängermodell
- 7) Subektiv niedrigste Windgeräusche, Mitbewerber um 22 % höher
- 8) Objektiv niedrigste Handling-Geräusche, Mitbewerber um 19 dB höher
- 9) Objektiv beste Impulsgeräusch-Absenkung, mindestens 10 dB besser als Mitbewerber



Bensheim Forum
14. Juni 2023

DNN-Training

Wie ein Hörsystem hören lernt

Wieviele Menschen braucht das Digitale?

