Expected Switching Time: a Markov Chain Based Performance Metric to Evaluate Auditory Attention Decoding Algorithms

SITB 2019

Simon Geirnaert, Tom Francart, and Alexander Bertrand

Hearing aids increase quality of life of the hearing impaired

1 out of 20
now suffers from
hearing loss

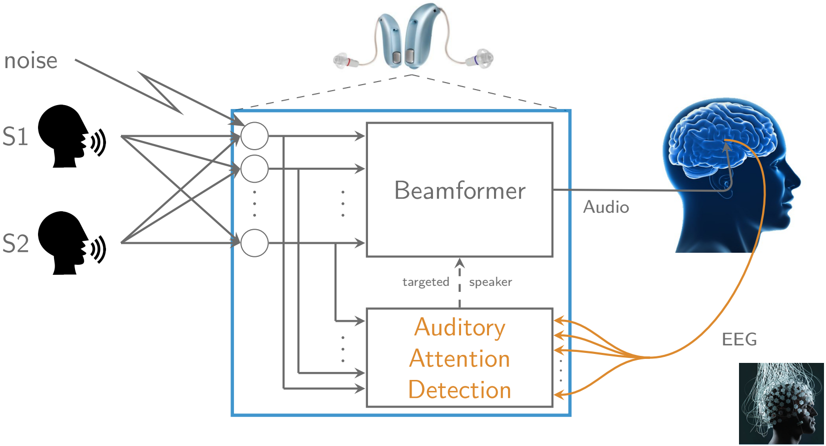
Estimation: 1 out of 10
in 2050

1.1 billion
young people at
risk of hearing loss

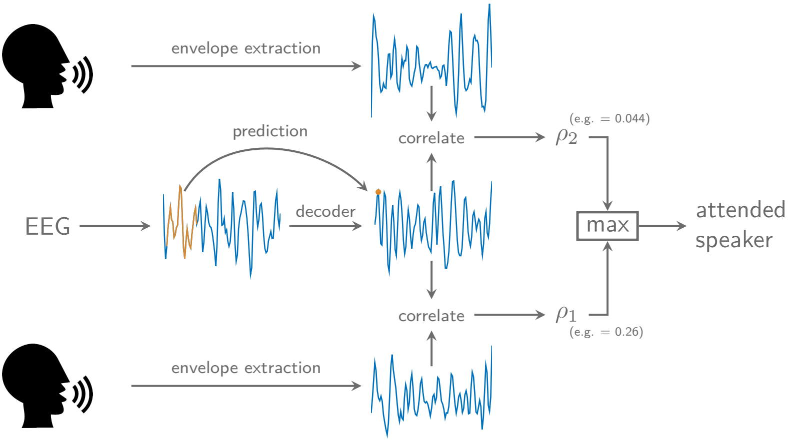
Current hearing aids:

- ✓ contain well-performing noise suppression algorithms (e.g., MWF)
- ✗ but lack information on the targeted speaker in a ‘cocktail party’ scenario

Neuro-steered hearing prostheses are the future

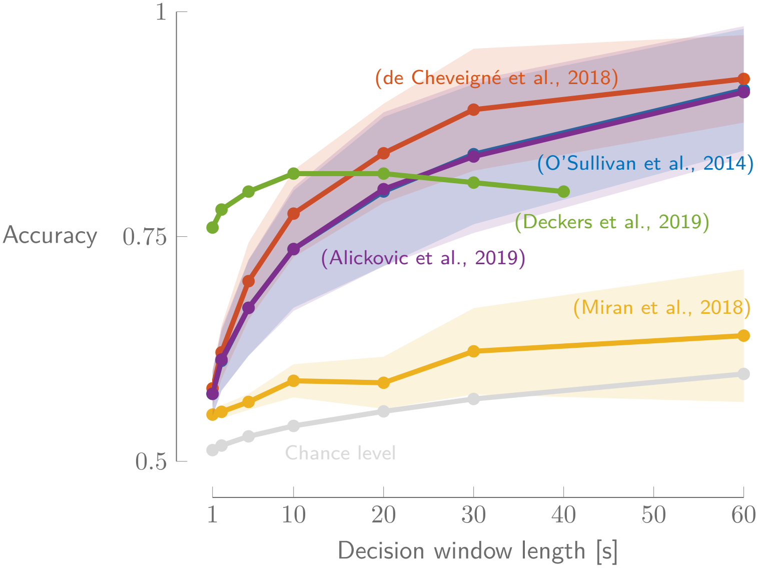


The framework of AAD algorithms



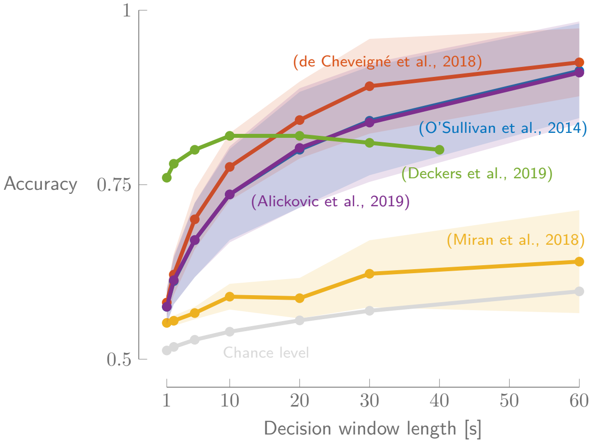
AAD = auditory attention detection

The $p(\tau)$-performance curves of AAD algorithms



However, AAD algorithms are evaluated in literature in a non-standardized fashion

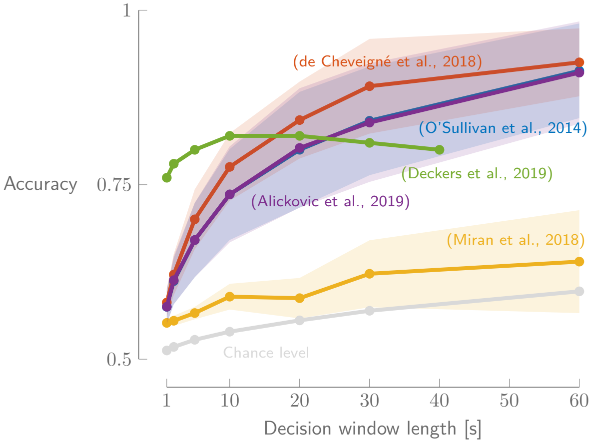
1 Multiple decision window lengths and accuracies represent the performance ...



However, AAD algorithms are evaluated in literature in a non-standardized fashion

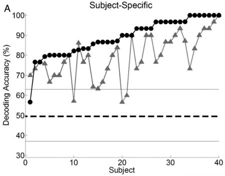
1 Multiple decision window lengths and accuracies represent the performance ...

2 ... which leads to potential inconclusiveness



However, AAD algorithms are evaluated in literature in a non-standardized fashion

1. Multiple decision window lengths and accuracies represent the performance ...
2. ...which leads to potential inconclusiveness
3. Different choice of decision window length obstructs fair comparison



(O'Sullivan et al., 2014), 60 s

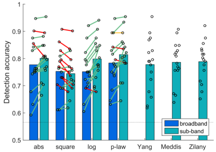


Fig. 2: Mean (bars) and individual subject (circles) detection accuracies for each of the different envelope extraction methods for a trial length of 30s. The dotted black line at 57% indicates the subject-specific detection accuracy which is only 5% likely to be surpassed by chance, based on a binomial distribution (success rate = 0.5, number of trials = 144).

(Biesmans et al., 2017), 30 s

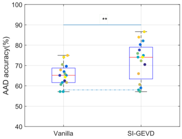
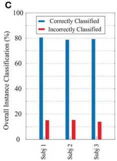


Fig. 5: Attention decoding performance on 20 s trials: SI-GEVD based forward modeling compared to forward modeling with mean of correlations over all channels ('vanilla'). The dotted line indicates chance level (95 percentile). Each box contains one mean decoding accuracy per subject, also indicated by colored points. In the plot, comparisons between methods are done using Wilcoxon's signed-rank tests: '**' for $p < 0.0100$.

(Das et al., 2019), 20 s



(Miran et al., 2018), 53 s

In search of a suitable performance metric: the requirements

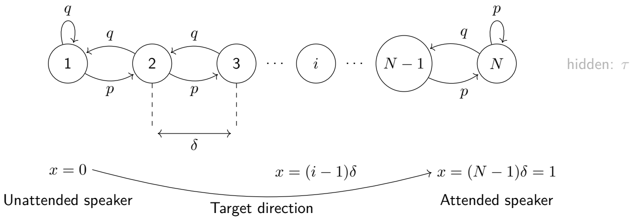
A performance metric for AAD algorithms should be:

1. single-number (allows for ranking, statistics, ...)
2. interpretable
3. combining accuracy and decision time
4. independent of evaluated decision window lengths

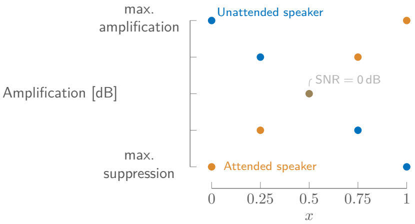
which motivates the design of a new metric: the

expected switching time (EST)

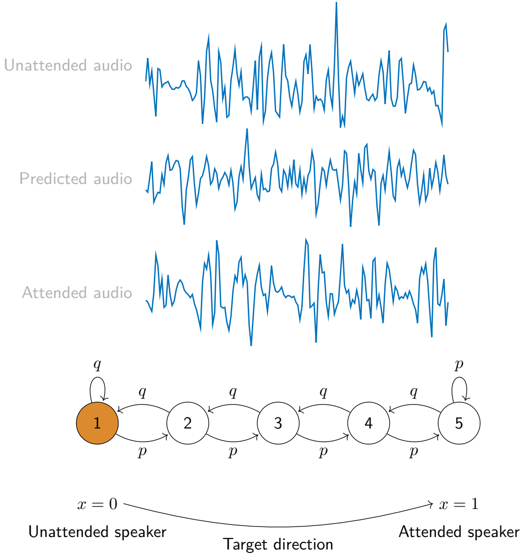
Modeling an adaptive gain control system as a Markov chain is crucial



Known parameters from the AAD algorithm: p, τ
Free design parameter: N (smoothness vs switching speed)



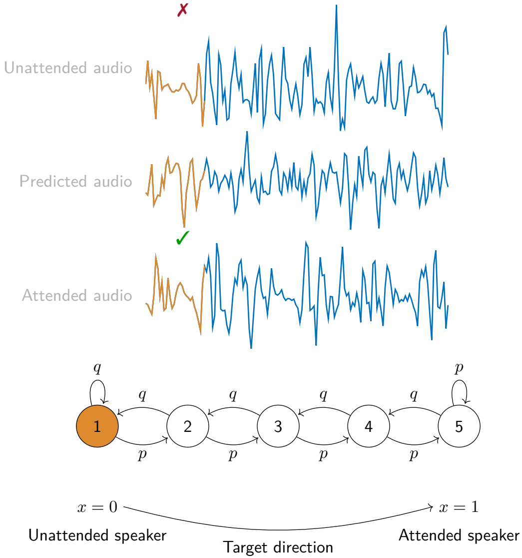
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \rightarrow 4$
Path	1
Switching time	

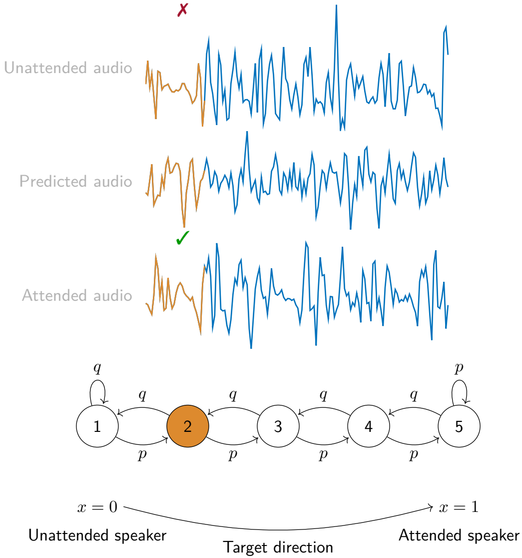
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \rightarrow 4$
Path	1
Switching time	

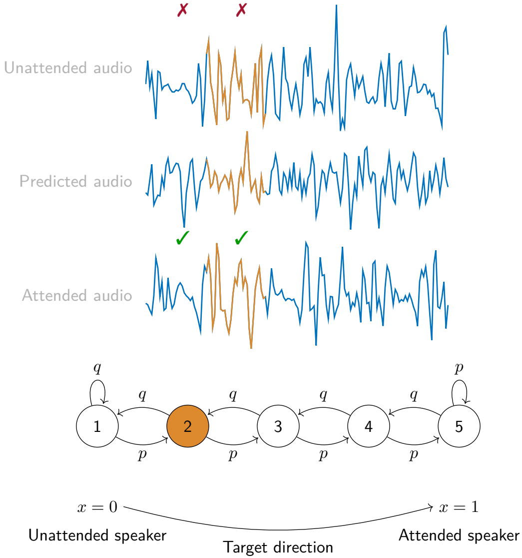
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \to 4$
Path	$1, 2$
Switching time	

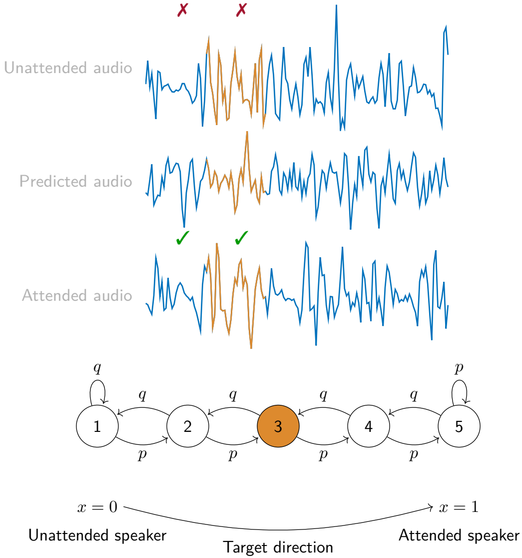
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \rightarrow 4$
Path	$1, 2$
Switching time	

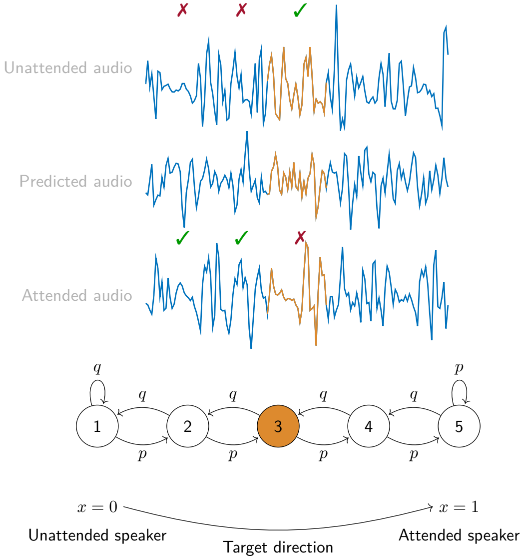
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \to 4$
Path	$1, 2, 3$
Switching time	

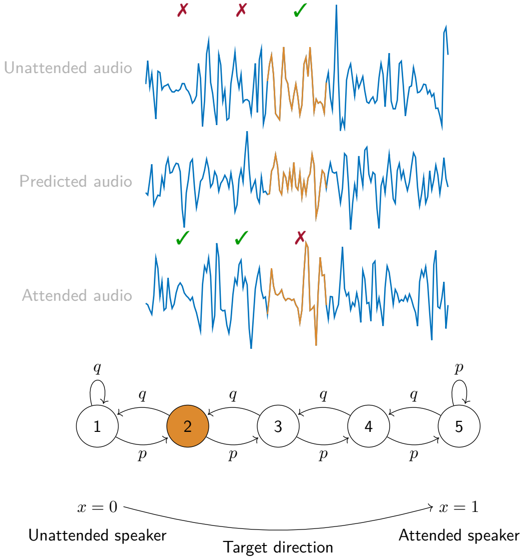
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \rightarrow 4$
Path	$1, 2, 3$
Switching time	

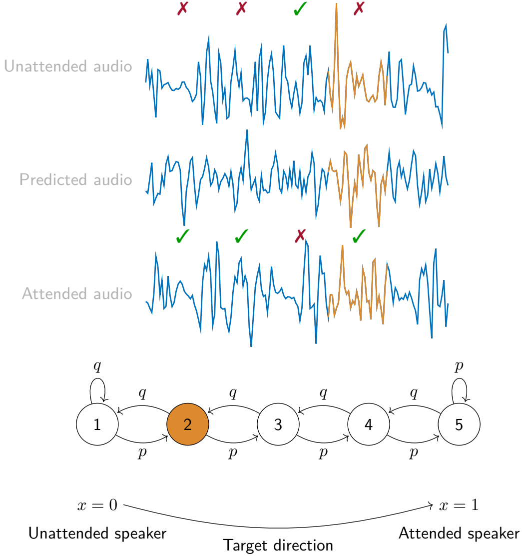
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \to 4$
Path	$1, 2, 3, 2$
Switching time	

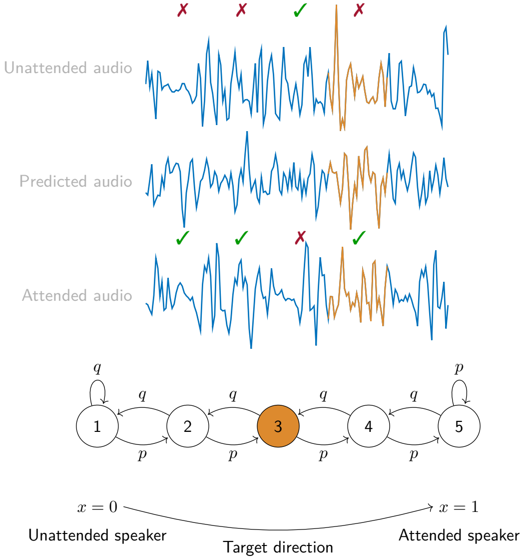
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \to 4$
Path	$1, 2, 3, 2$
Switching time	

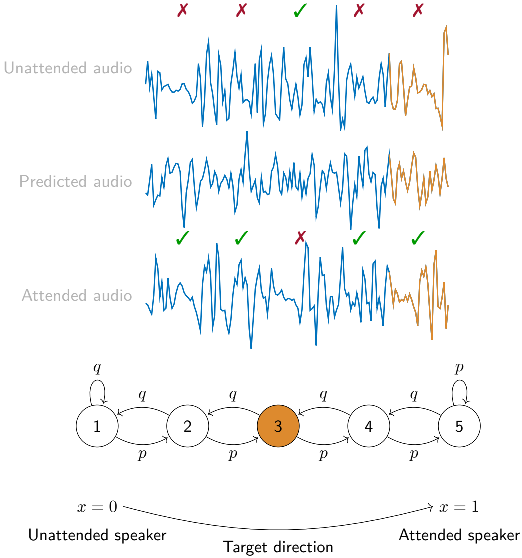
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \rightarrow 4$
Path	$1, 2, 3, 2, 3$
Switching time	

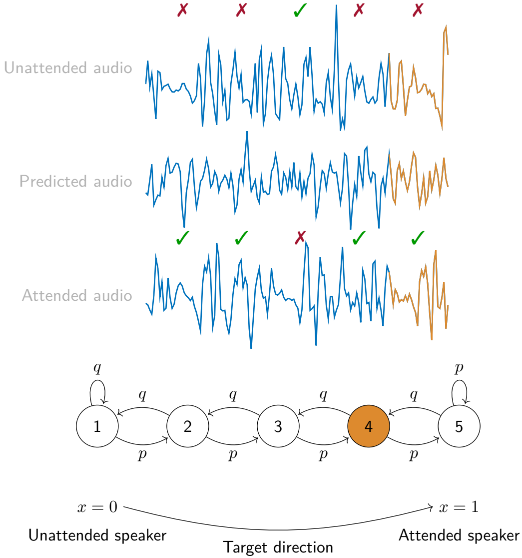
The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \rightarrow 4$
Path	$1, 2, 3, 2, 3$
Switching time	

The Markov chain as adaptive gain control system: an example



AAD algorithm
with $p = 80\%$
for 5 s windows

Attention switch	$1 \rightarrow 4$
Path	$1, 2, 3, 2, 3, 4$
Switching time	$5 \times 5\,\text{s} = 25\,\text{s}$

The design of the expected switching time consists of several subproblems

Expected switching time: definition in words

The expected switching time (EST) is the expected time required to reach the P_0-confidence interval, containing the comfortable level c, after an attention switch, in an optimized Markov chain as a model for an adaptive gain control system in a neuro-steered hearing prosthesis.

The design of the expected switching time consists of several subproblems

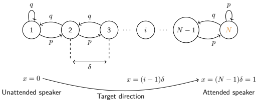
Expected switching time (EST)

optimal Markov chain

switching time

optimal Markov chain

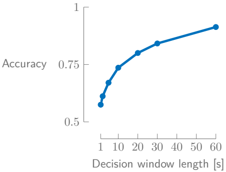
Optimal number of states N



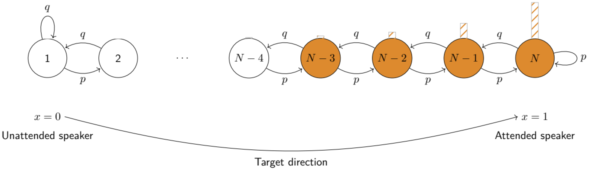
Quantification of switching time in Markov chain

Transit time: $T(p(\tau), \tau, N)$

Optimal working point (τ, p) on performance curve



The P_0-confidence interval of the Markov chain model



The P_0-confidence interval can be found by searching for the largest $\bar{k}$ such that:

$$\sum_{j=\bar{k}}^{N} \pi(j) \geq P_0 \Leftrightarrow \bar{k} = \left\lfloor \frac{\log\left(r^N(1-P_0)+P_0\right)}{\log(r)} + 1 \right\rfloor,$$

using

$$r = \frac{p}{1-p} \text{ and steady-state distribution } \pi(i) = \frac{r-1}{r^N-1} r^{i-1}$$

The P_0-confidence interval of the Markov chain model

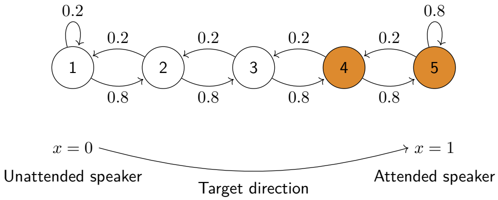
The P_0-confidence interval

$$[\bar{x}, 1] = \left[\frac{\bar{k}-1}{N-1}, 1\right],$$

with

$$\bar{k} = \left\lfloor \frac{\log\left(r^N(1-P_0)+P_0\right)}{\log(r)} + 1 \right\rfloor,$$

Example: the 90%-confidence interval of this Markov chain



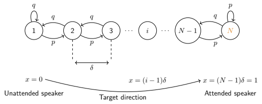
is $[0.75, 1]$ ($\bar{k} = 4$)

Design constraints on the Markov chain: optimizing free design parameter N

Expected switching time (EST)

optimal Markov chain · switching time · optimal Markov chain

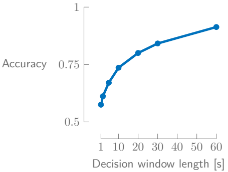
Optimal number of states N



Quantification of switching time in Markov chain

Transit time: $T(p(\tau), \tau, N)$

Optimal working point (τ, p) on performance curve



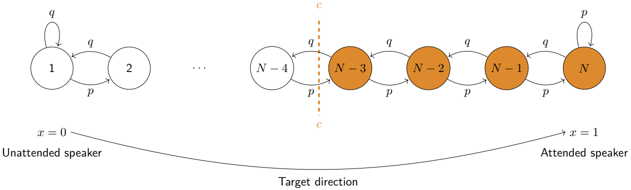
Design constraints on the Markov chain: optimizing free design parameter N

Given (τ, p)-performance, two goals:

1. Optimize free design parameter N
2. Interpret model in hearing aid context

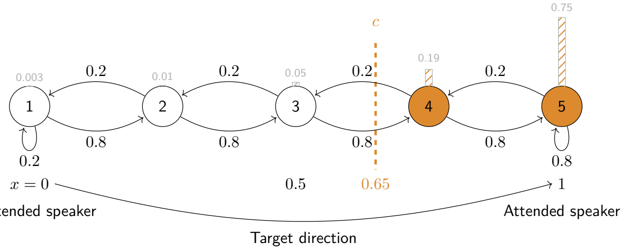
result in two design criteria:

1. $\bar{x} \in [c, 1]$, with c a predefined desired lower bound of the P_0-confidence interval
2. $N \geq N_{\text{min}}$: obtain smooth transitions in the gain adaptation



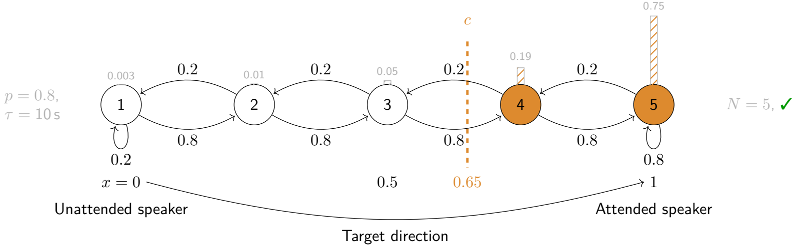
Design constraints on the Markov chain: optimizing free design parameter N

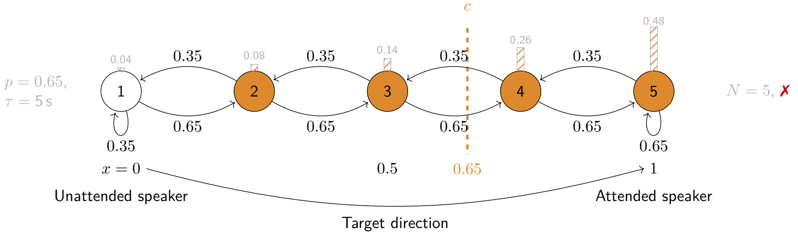
$p = 0.8$,
$\tau = 10\,\text{s}$



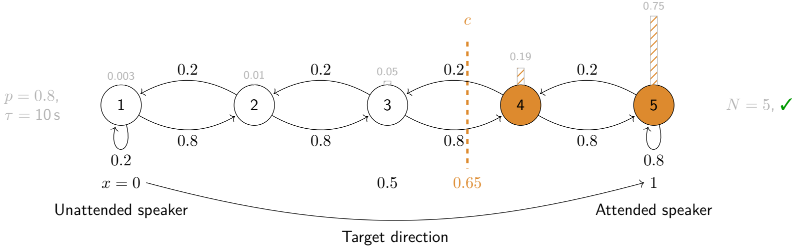
$N = 5$, ✓

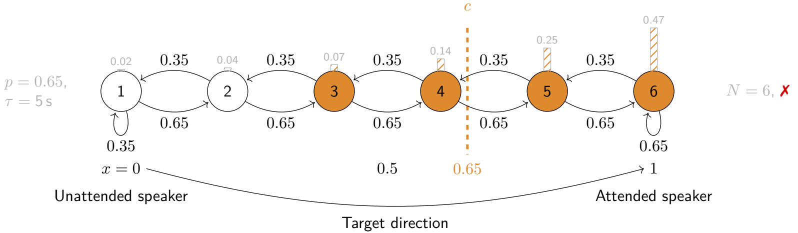
Design constraints on the Markov chain: optimizing free design parameter N



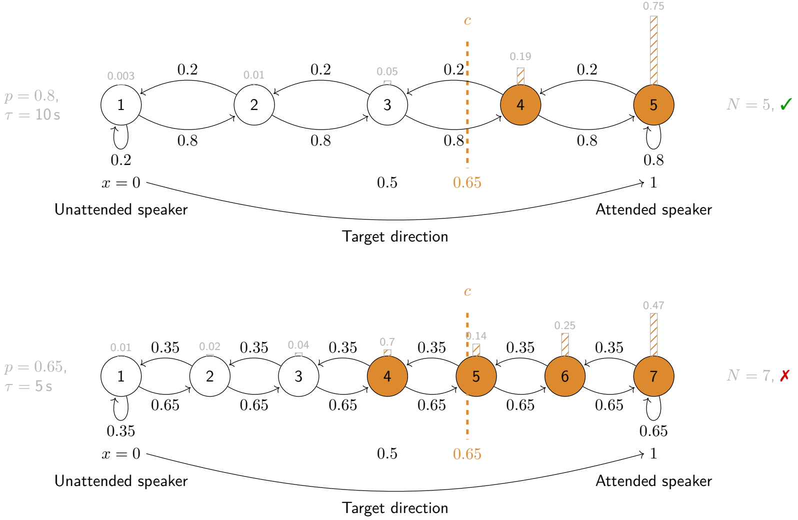


Design constraints on the Markov chain: optimizing free design parameter N

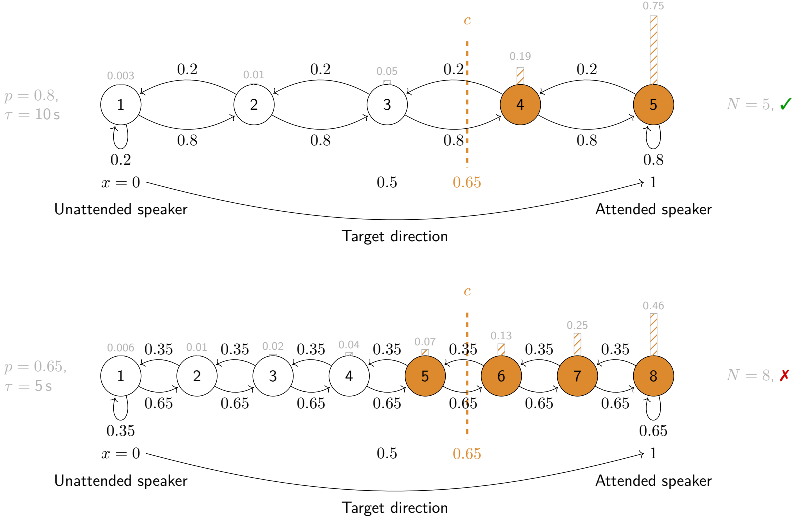




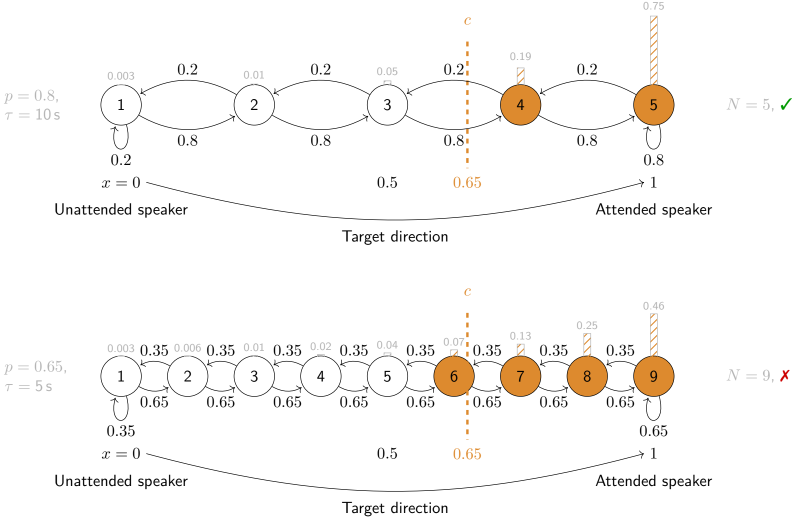
Design constraints on the Markov chain: optimizing free design parameter N



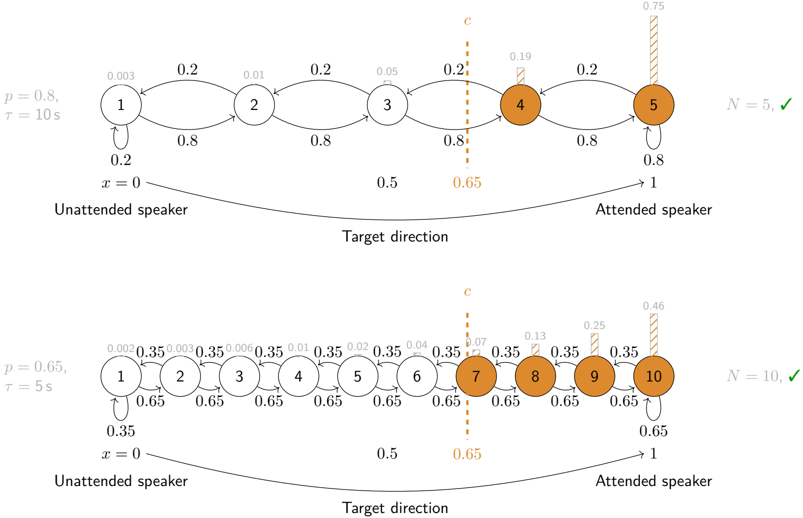
Design constraints on the Markov chain: optimizing free design parameter N



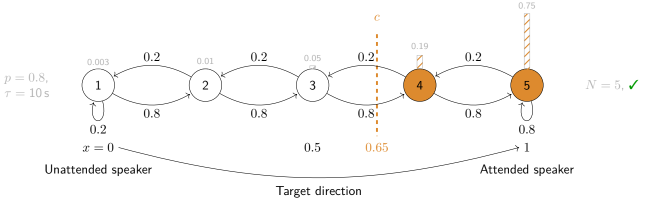
Design constraints on the Markov chain: optimizing free design parameter N



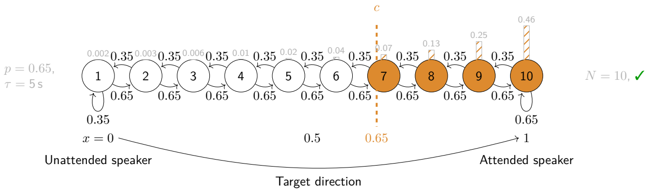
Design constraints on the Markov chain: optimizing free design parameter N



Design constraints on the Markov chain: optimizing free design parameter N



$\frac{p}{\tau}\left[\frac{\#\ \text{correct decisions}}{\text{s}}\right]$	0.08
EST [s]	40.81



$\frac{p}{\tau}\left[\frac{\#\ \text{correct decisions}}{\text{s}}\right]$	0.13
EST [s]	70.58

The transit time: the expected time needed to arrive in target state $\bar{k}$, starting from *any* initial state $i < \bar{k}$

Expected switching time (EST)

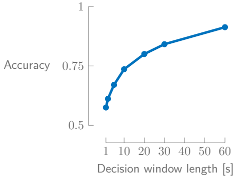
optimal Markov chain ← **Optimal number of states N**

switching time ← **Quantification of switching time in Markov chain**

optimal Markov chain ← **Optimal working point (τ, p) on performance curve**

1. $\bar{x} \in [c, 1]$
2. $N \geq N_{\text{min}}$

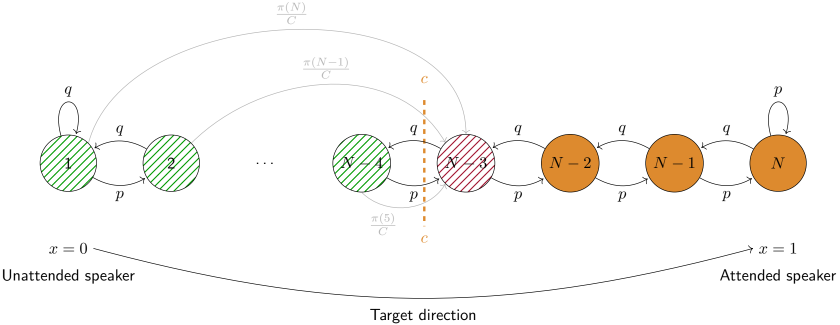
Transit time: $T(p(\tau), \tau, N)$



The transit time: the expected time needed to arrive in target state $\bar{k}$, starting from *any* initial state $i < \bar{k}$

An attention switch

An attention switch is defined as the transition from *any* initial state $i < \bar{k}$ outside the P_0-confidence interval to lower bound $\bar{k}$.



The transit time: the expected time needed to arrive in target state $\bar{k}$, starting from *any* initial state $i < \bar{k}$

An attention switch

An attention switch is defined as the transition from *any* initial state $i < \bar{k}$ outside the P_0-confidence interval to lower bound $\bar{k}$.

We call the associated switching time the transit time:

$$T(p(\tau), \tau, N) = \tau \mathbb{E}\{s | i \to \bar{k}, \forall\, i < \bar{k}\} = \tau \sum_{s=0}^{+\infty} s P(s | i \to \bar{k}, \forall\, i < \bar{k})$$

$$= \tau \frac{r^{\bar{k}+1} - r^{\bar{k}}}{r^{\bar{k}} - r} \sum_{i=1}^{\bar{k}-1} r^{-i} h_{\bar{k}}(i),$$

with mean hitting time $h_j(i) = \mathbb{E}\{s | i \to j\} = \frac{j-i}{2p-1} + \frac{p(r^{-j} - r^{-i})}{(2p-1)^2}, \forall\, i \leq j$

Optimizing the transit time over the performance curve results in the EST (bis)

Expected switching time (EST)

optimal Markov chain

switching time

optimal Markov chain

Optimal number of states N

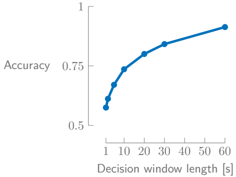
Quantification of switching time in Markov chain

Optimal working point (τ, p) on performance curve

1. $\bar{x} \in [c, 1]$
2. $N \geq N_{\text{min}}$

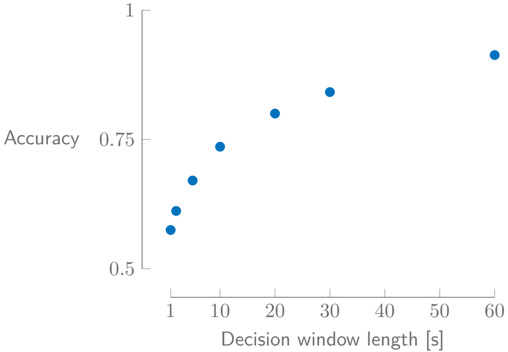
Transit time:

$$T(p(\tau), \tau, N) = \tau \frac{r^{\bar{k}+1} - r^{\bar{k}}}{r^{\bar{k}} - r} \sum_{i=1}^{\bar{k}-1} r^{-i} h_{\bar{k}}(i)$$



Optimizing the transit time over the performance curve results in the EST (bis)

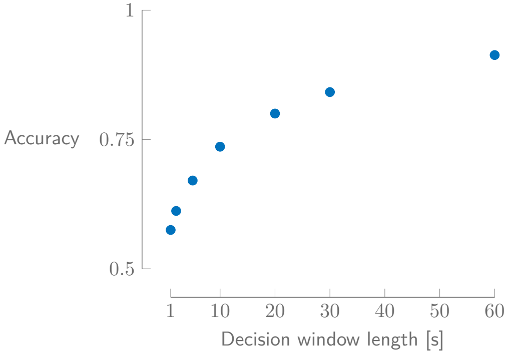
? At which accuracy p and decision window length τ should the Markov chain be optimized and the transit time computed?



Optimizing the transit time over the performance curve results in the EST (bis)

? At which accuracy p and decision window length τ should the Markov chain be optimized and the transit time computed?

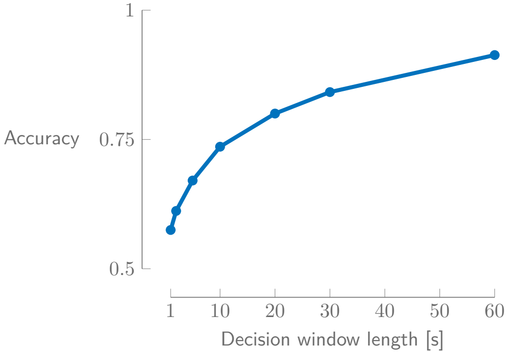
1 Construct the $p(\tau)$-performance curve via linear interpolation



Optimizing the transit time over the performance curve results in the EST (bis)

? At which accuracy p and decision window length τ should the Markov chain be optimized and the transit time computed?

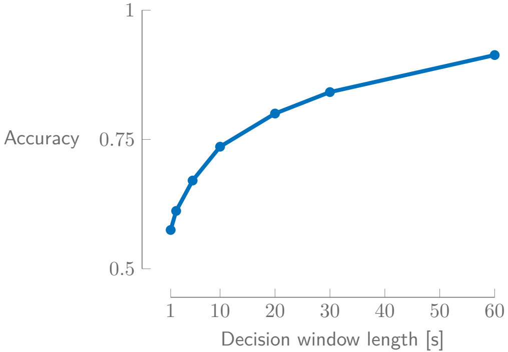
1. Construct the $p(\tau)$-performance curve via linear interpolation



Optimizing the transit time over the performance curve results in the EST (bis)

? At which accuracy p and decision window length τ should the Markov chain be optimized and the transit time computed?

1. Construct the $p(\tau)$-performance curve via linear interpolation
2. Sample the $p(\tau)$-performance curve

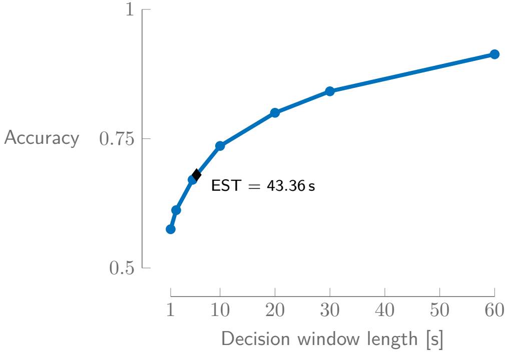


Optimizing the transit time over the performance curve results in the EST (bis)

? At which accuracy p and decision window length τ should the Markov chain be optimized and the transit time computed?

1. Construct the $p(\tau)$-performance curve via linear interpolation
2. Sample the $p(\tau)$-performance curve
3. Pick the working point with the lowest transit time

leads to the expected switching time



Bringing it all together: the expected switching time

Expected switching time (EST)

optimal Markov chain — switching time — optimal Markov chain

Optimal number of states N

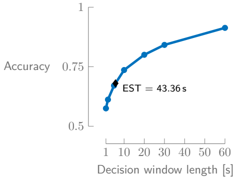
1. $\bar{x} \in [c, 1]$
2. $N \geq N_{\text{min}}$

Quantification of switching time in Markov chain

Transit time:

$$T(p(\tau), \tau, N) = \tau \frac{r^{\bar{k}+1} - r^{\bar{k}}}{r^{\bar{k}} - r} \sum_{i=1}^{\bar{k}-1} r^{-i} h_{\bar{k}}(i)$$

Optimal working point (τ, p) on performance curve



Bringing it all together: the expected switching time

Expected switching time: full definition

The expected switching time (EST) is the expected time required to reach the P_0-confidence interval[1], containing the comfortable level c, after an attention switch, in an optimized Markov chain as a model for an adaptive gain control system in a neuro-steered hearing prosthesis. It is the solution of the following optimization problem:

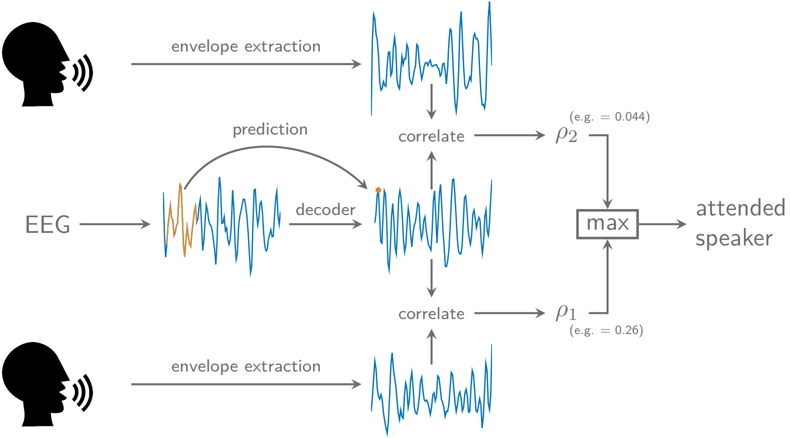
$$\begin{aligned} \text{EST} = \quad & \min_{N,\tau} \quad T(p(\tau), \tau, N) \\ & \text{s.t.} \quad \bar{x} \in [c, 1] \\ & \qquad\quad N \geq N_{\text{min}} \end{aligned}$$

where $T(p(\tau), \tau, N)$ is the transit time and $\bar{x} = \frac{\bar{k}-1}{N-1}$, with $\bar{k}$ the lower bound of the P_0-confidence interval.

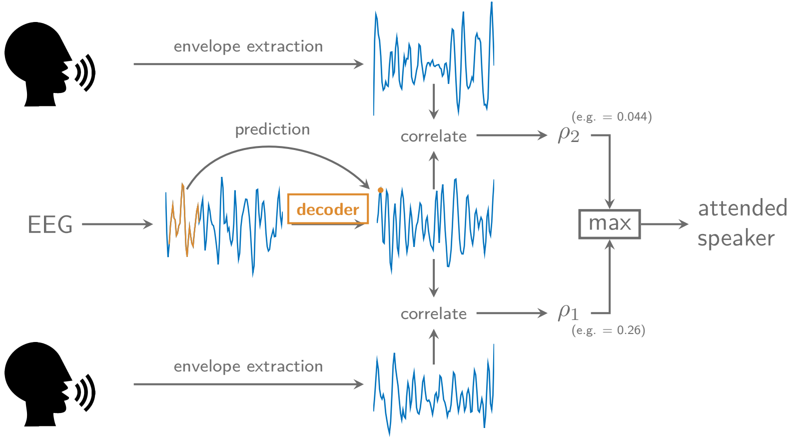
The computation of the EST is easy given that $T(p(\tau), \tau, N)$ is monotonically nondecreasing with N

[1] Starting from a stochastically defined initial state based on the steady-state distribution

The framework of AAD algorithms



The framework of AAD algorithms



MMSE as example of backward AAD decoder

Basic approach via MMSE: (O'Sullivan et al., 2014)

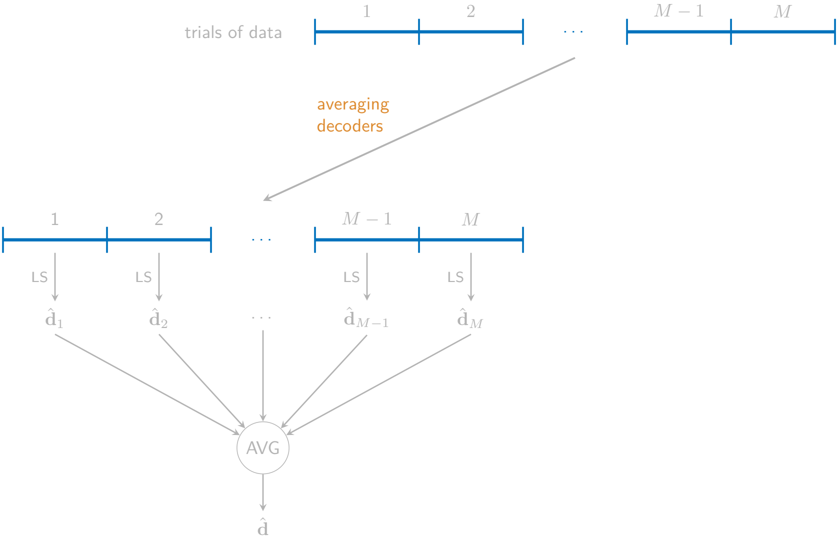
$$\min_{d_n(l)} E\{(s_{\mathsf{att}}(t) - \sum_{n=1}^{N} \sum_{l=0}^{L-1} d_n(l) x_n(t+l))^2\},$$

with N = number of channels (64) and L = length integration window (250 ms). Vectorized solution:

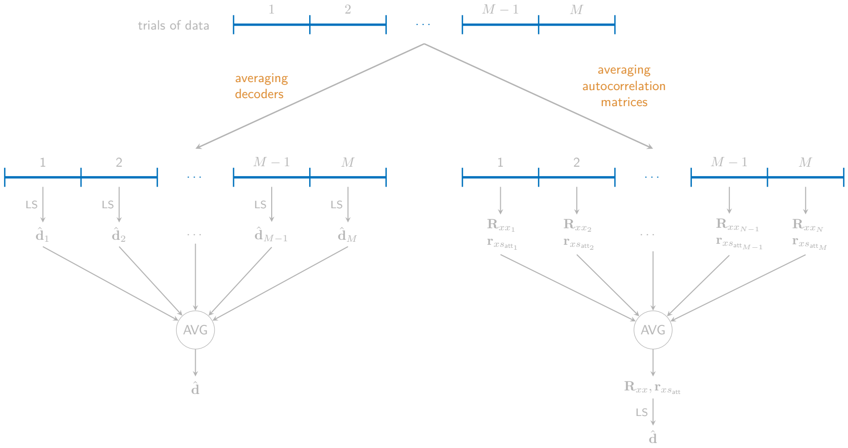
$$\boxed{\hat{\mathbf{d}} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{x s_{\mathsf{att}}}}$$

MMSE as example of backward AAD decoder

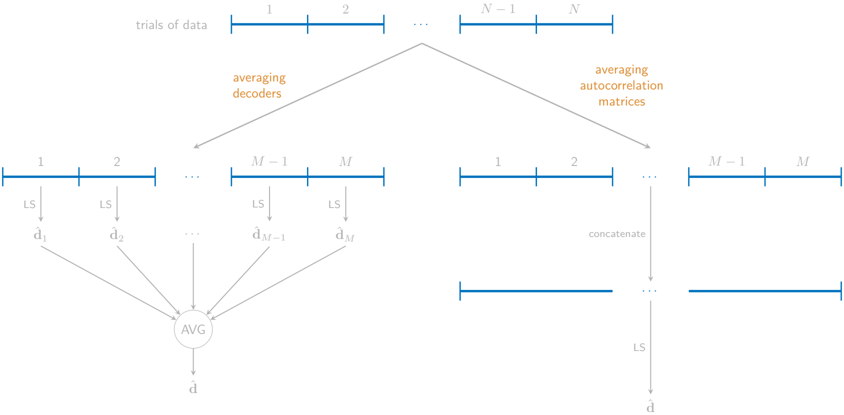
MMSE as example of backward AAD decoder



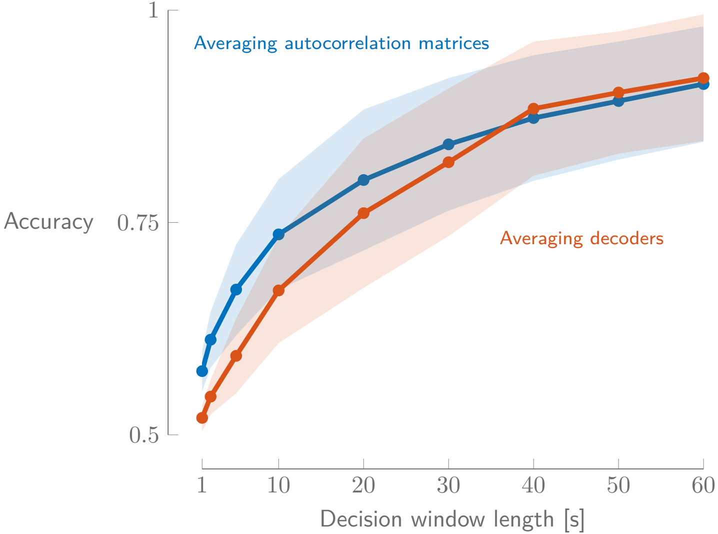
MMSE as example of backward AAD decoder



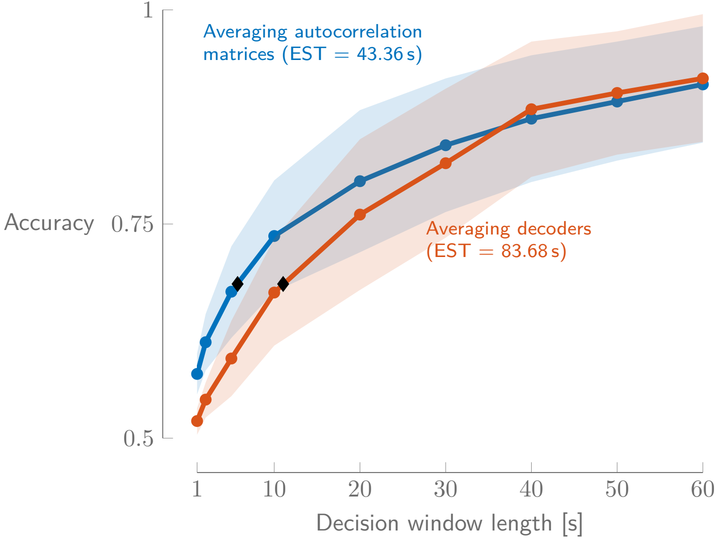
MMSE as example of backward AAD decoder



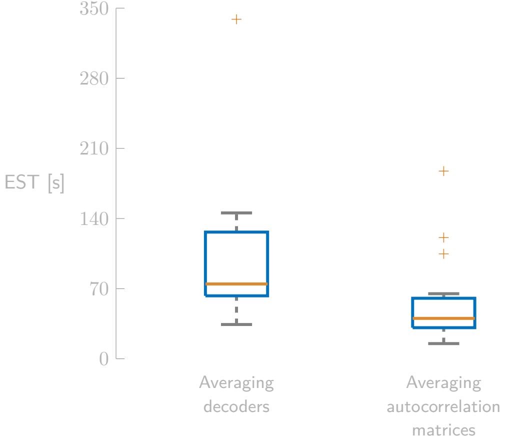
Averaging decoders vs autocorrelation matrices: what does the EST decide?



Averaging decoders vs autocorrelation matrices: what does the EST decide?



Averaging decoders vs autocorrelation matrices: what does the EST decide?



A paired, one-sided Wilcoxon signed rank test shows that averaging autocorrelation matrices > averaging decoders

Interested in using the EST metric?

https://github.com/exporl/est-toolbox

Questions? Remarks? Thoughts? Ideas?

simon.geirnaert@esat.kuleuven.be

KU LEUVEN

STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics

KU LEUVEN expORL

fwo Research Foundation Flanders Opening new horizons

Hyperparameter choice: N_{min}, P_0 and c

$$\boxed{\begin{aligned} \text{EST} = \quad & \min_{N,\tau} \quad T(p(\tau),\tau,N) \\ & \text{s.t.} \quad \bar{x} \in [c,1] \\ & \qquad\;\; N \geq N_{\text{min}} \end{aligned}}$$

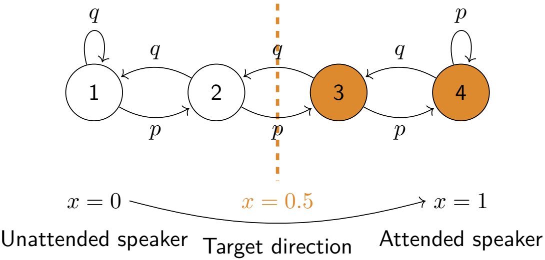
Three parameters involved in the design constraints:

Hyperparameter choice: N_{min}, P_0 and c

$$\begin{aligned} \text{EST} = \quad & \min_{N,\tau} \quad T(p(\tau), \tau, N) \\ & \text{s.t.} \quad \bar{x} \in [c, 1] \\ & \qquad\ \ N \geq N_{\text{min}} \end{aligned}$$

Three parameters involved in the design constraints:

- $N_{\text{min}} = 5$



Assume $p < P_0$

Hyperparameter choice: N_{min}, P_0 and c

$$\begin{aligned} \mathsf{EST} = \quad & \min_{N,\tau} \quad T(p(\tau), \tau, N) \\ & \text{s.t.} \quad \bar{x} \in [c, 1] \\ & \qquad\; N \geq N_{\mathsf{min}} \end{aligned}$$

Three parameters involved in the design constraints:

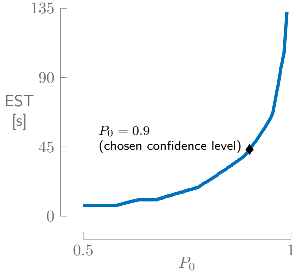
- $N_{\mathsf{min}} = 5$
- $c = 0.65$

Hyperparameter choice: N_{min}, P_0 and c

$$\begin{aligned} \text{EST} = \quad & \min_{N,\tau} \quad T(p(\tau), \tau, N) \\ & \text{s.t.} \quad \bar{x} \in [c, 1] \\ & \qquad\quad N \geq N_{\text{min}} \end{aligned}$$

Three parameters involved in the design constraints:

- $N_{\text{min}} = 5$
- $c = 0.65$
- $P_0 = 0.9$



Average over O'Sullivan et al., $c = 0.65$

The EST in case dependent decisions due to overlapping decision windows

Independence condition in Markov chain is violated when there is overlap in decision windows.
Two questions:

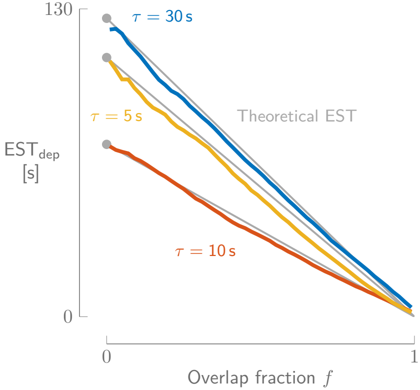
1. What is the difference between EST_{dep} and theoretical EST?
2. Is the theoretical EST still a valid *relative* performance metric?

The EST in case dependent decisions due to overlapping decision windows

Independence condition in Markov chain is violated when there is overlap in decision windows.

Two questions:

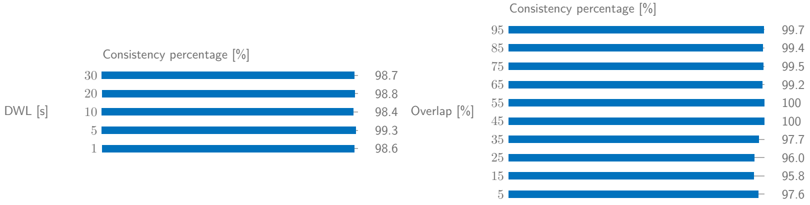
1. What is the difference between EST_{dep} and theoretical EST? Acceptable
2. Is the theoretical EST still a valid *relative* performance metric?



The EST in case dependent decisions due to overlapping decision windows

Independence condition in Markov chain is violated when there is overlap in decision windows.
Two questions:

1. What is the difference between EST_{dep} and theoretical EST? Acceptable
2. Is the theoretical EST still a valid *relative* performance metric? Yes



Average consistency $= 98.73\%$

The algorithm to compute the EST

Algorithm 1 Computation of the EST metric

Input: Evaluated points on the $p(\tau)$-performance curve $(\tau_i, p_i), i \in \{1, \dots, I\}$, the required number of interpolated samples K of the performance curve $p(\tau)$ and the hyperparameters: confidence interval P_0, lower bound c and minimum number of states N_{min}. The suggested default values are $K = 1000, P_0 = 0.9, c = 0.65$ and $N_{\text{min}} = 5$ (see Section III-A).

Output: EST

1: Construct K samples of the performance curve $p(\tau)$ by piecewise linear interpolating through evaluated points $(\tau_i, p_i), i \in \{1, \dots, I\}$

2: **for** each sampled τ **do**

3: Find $\hat{N}_\tau$ by going over the candidate values $N = N_{\text{min}} + i$, with $i = 0, 1, 2, \dots,$ in this specific order, until the first value N is found that satisfies:

$$\frac{\bar{k}-1}{N-1} \geq c, \quad \text{and } N \geq N_{\text{min}},$$

with $\bar{k} = \left\lfloor \frac{\log(r^N(1-P_0)+P_0)}{\log(r)} + 1 \right\rfloor$ and $r = \frac{p(\tau)}{1-p(\tau)}$.

4: Given $\hat{N}_\tau$, compute the transit time $T(p(\tau), \tau, \hat{N}_\tau) = \tau \frac{r^{\bar{k}+1} - r^{\bar{k}}}{r^{\bar{k}} - r} \sum_{i=1}^{\bar{k}-1} r^{-i} h_{\bar{k}}(i)$, with $h_{\bar{k}}(i) = \frac{\bar{k}-i}{2p-1} + \frac{p(r^{-\bar{k}} - r^{-i})}{(2p-1)^2}$.

5: **end for**

6: The EST is equal to the minimum transit time over all sampled τ:

$$\text{EST} = \min_{\tau} \; T(p(\tau), \tau, \hat{N}_\tau).$$

Comparison between less interpretable ITR and EST

Information transfer rate (ITR $[\frac{\text{bit}}{\text{s}}]$) is defined as:

$$\text{ITR} = \frac{1}{\tau}\left(\log_2 M + p\log_2 p + (1-p)\log_2 \frac{1-p}{M-1}\right),$$

with $M = 2$ here.