Multi-episodic Perceived Quality of Telecommunication Services

Dennis Guse, M. Sc. | 2016-09-01

Outline

Perceived Quality

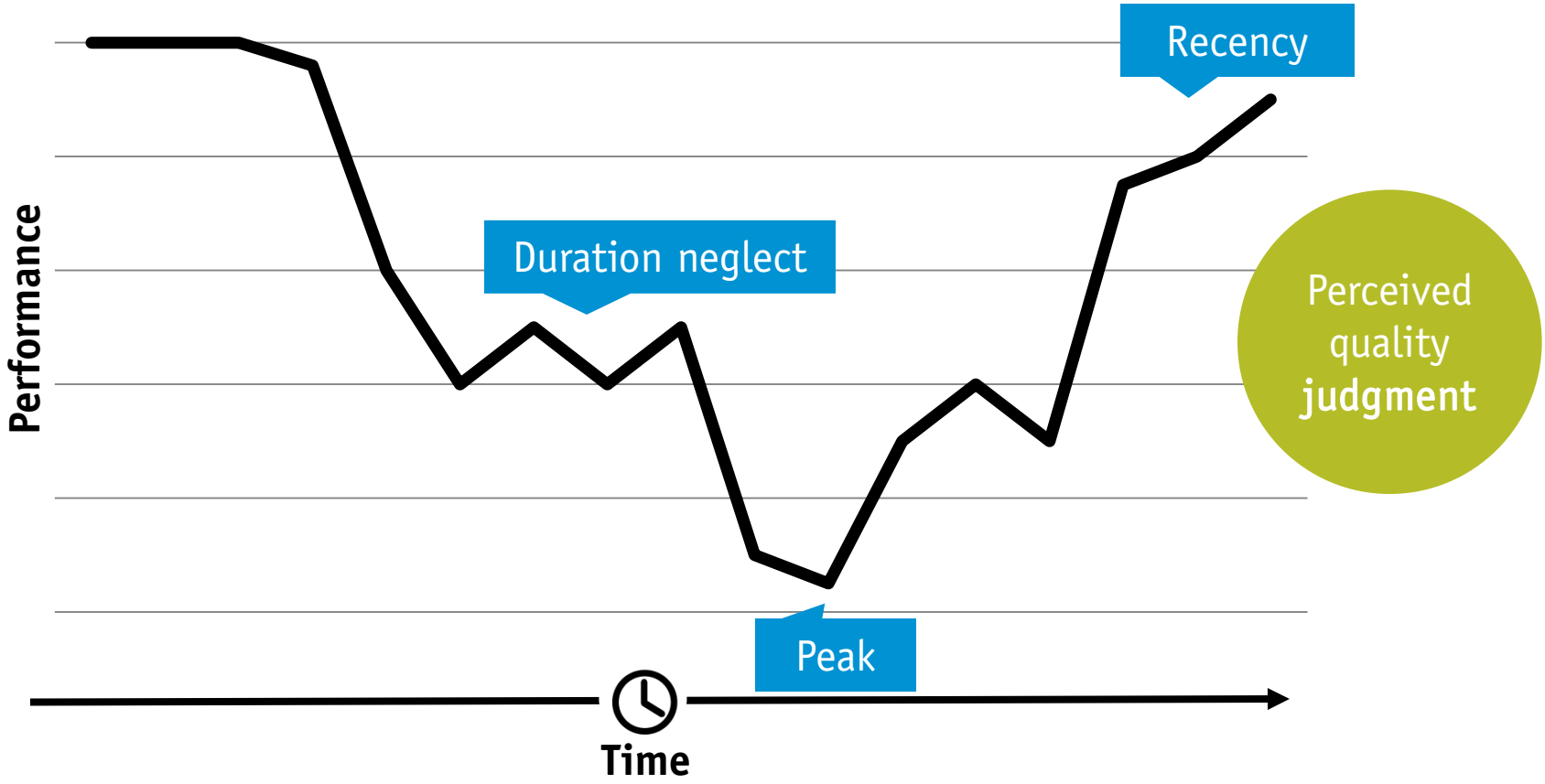
Influence factors

- Human
- System
- Context

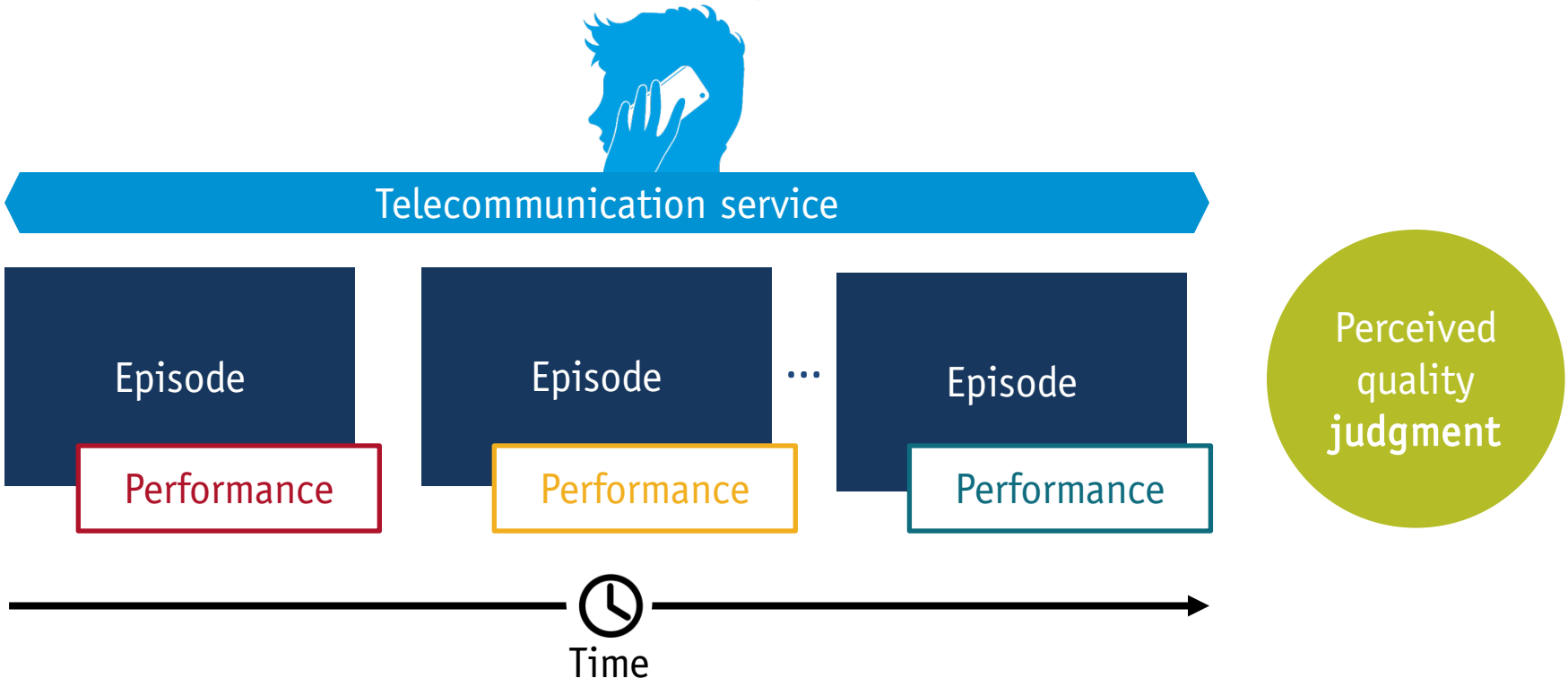
(Reiter et al., 2014)

Retrospective Judgments

Example of for one experience



Multi-episodic Perceived Quality



Research Question

Research Question:

How does *multi-episodic perceived quality* evolve over several, distinct usage episodes with a single telecommunication service?

Usage episode as …

» distinct, meaningful, and self-contained interaction by a user with a telecommunication service to achieve his goal(s) «

- based upon episodic memory

Research Goals

🎯 #1

- to investigate how the performance of *a sequence of usage episodes* determines judgments of multi-episodic perceived quality

🎯 #2

- to investigate how multi-episodic judgments can be *predicted*

The Defined-use Method

#1: Investigate the formation process of multi-episodic judgments

First applied by *Möller et al. (2011)*
- 12 days
- 5 Multi-episodic conditions
- Video telephony (Skype)

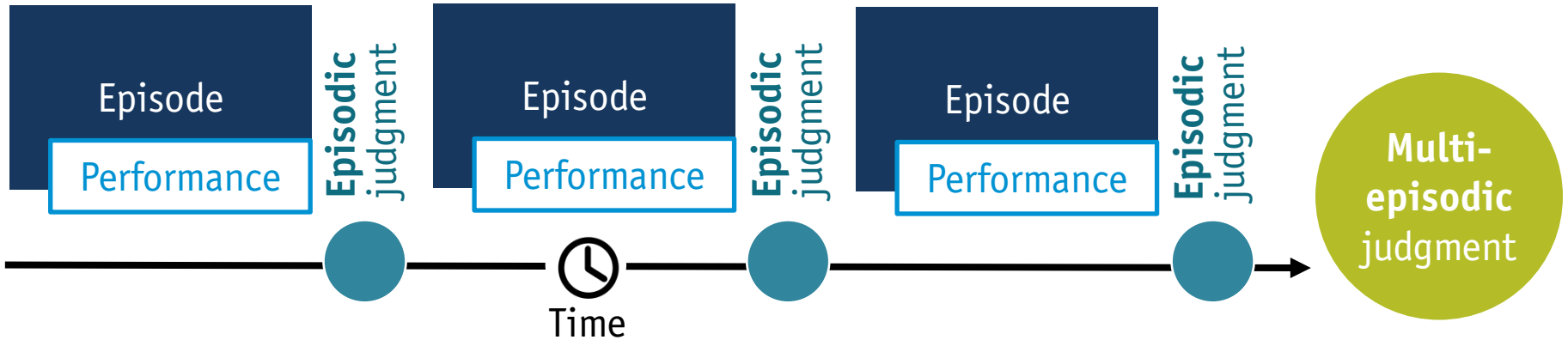
Defined-use method

➤ **create <u>multi-episodic conditions</u> by defining each usage episode:**
- when, how, and what
- performance

- repeatable conditions: enables to derive a *Mean Opinion Score* (MOS)

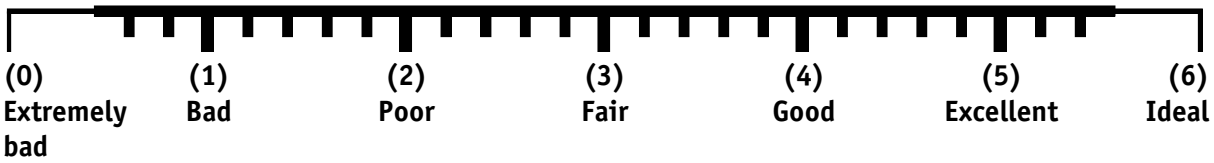
- unrealistic
- technically and organizationally challenging

Judgments



Episodic judgment: How do you judge the quality of the just finished usage episode?

Multi-episodic judgment: How do you judge the quality of all usage episodes so far?



Empirical Investigation

Applying the defined-use method:

- task-driven usage
- constant performance per usage episode
- first usage episodes in best performance
- **speech-only** service
- **new** service

Aspects

Usage periods

- continuous use (1h)
- distributed use (6 days)

Service types

- speech telephony
- audio book

User behavior

- conversation
- listening / consumption

Experimental Design

Usage Period: 1h

- 6 usage episodes
- multi-episodic judgment after the 3rd and 6th usage episode

E1: Two-party telephony (3-7min)

- task: conversation (ITU-T P.805)

E2a: Third-party listening (2-3min)

- task: consume recordings (ITU-T P.805)

E3: Audio book (3min)

- task: consume content

Usage Period: 6 days

- 2 usage episodes per day
- multi-episodic judgment after the 3rd and 6th day

E6: Audio book (12-17min)

- task: consume content

Performance levels: ▲ **High Performance** (HP): G.722 and CD
▼ **Low Performance** (LP): LPC-10

Empirical Results

Reported as **MOS**

Participants

E1: 129

E2a: 115

E3: 36

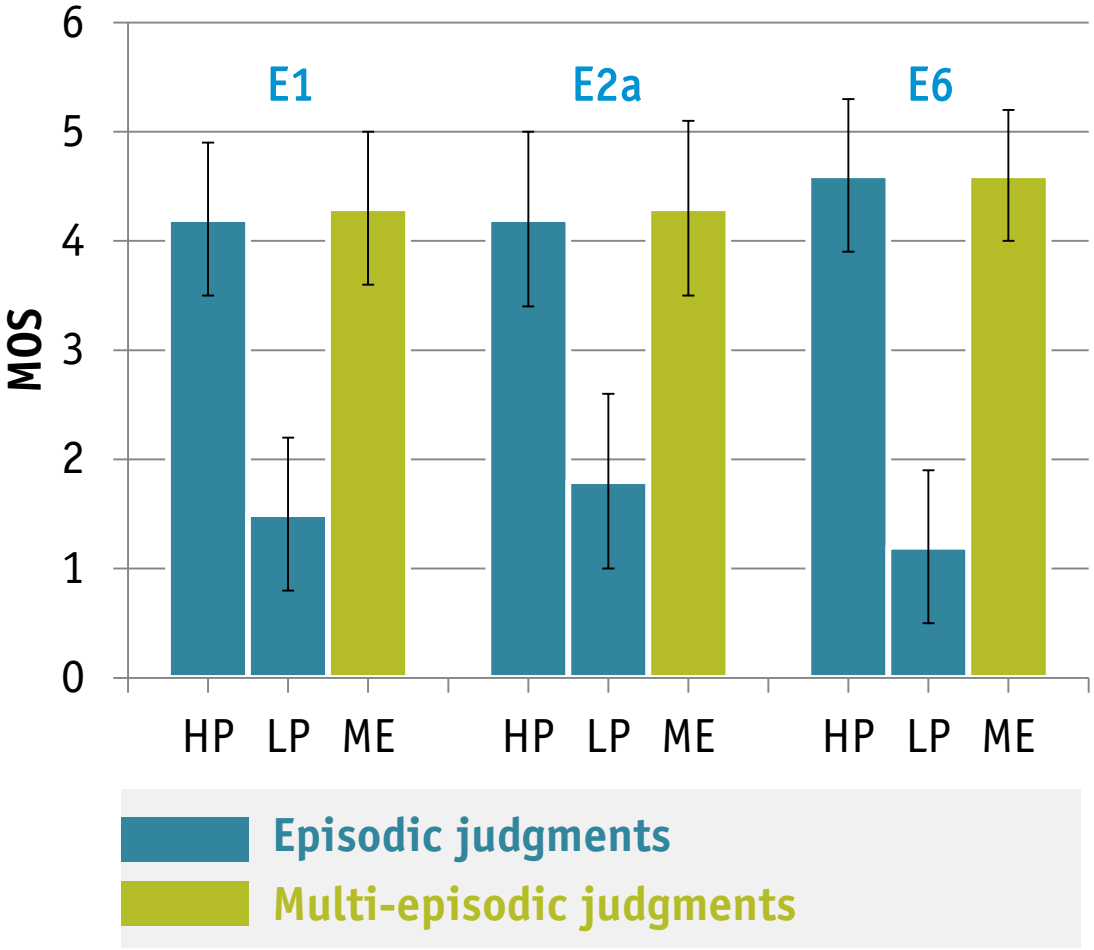
E6: 94

german speaking

younger than 35

mainly students

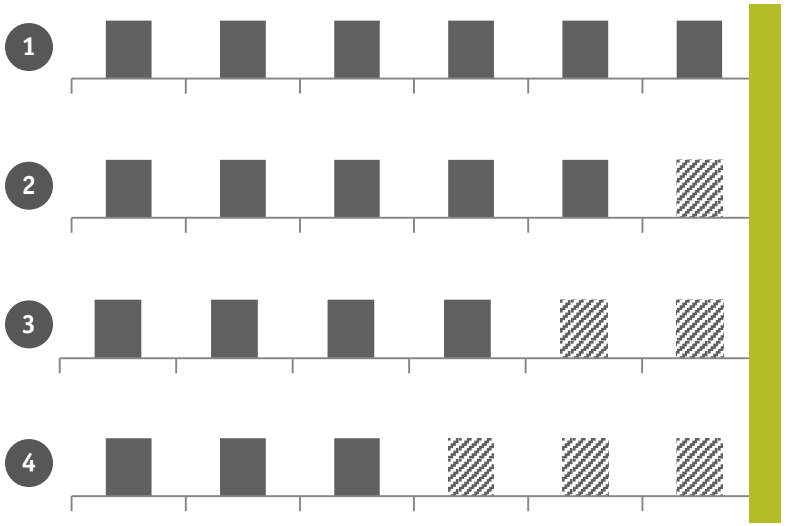
3 of 8 hypotheses



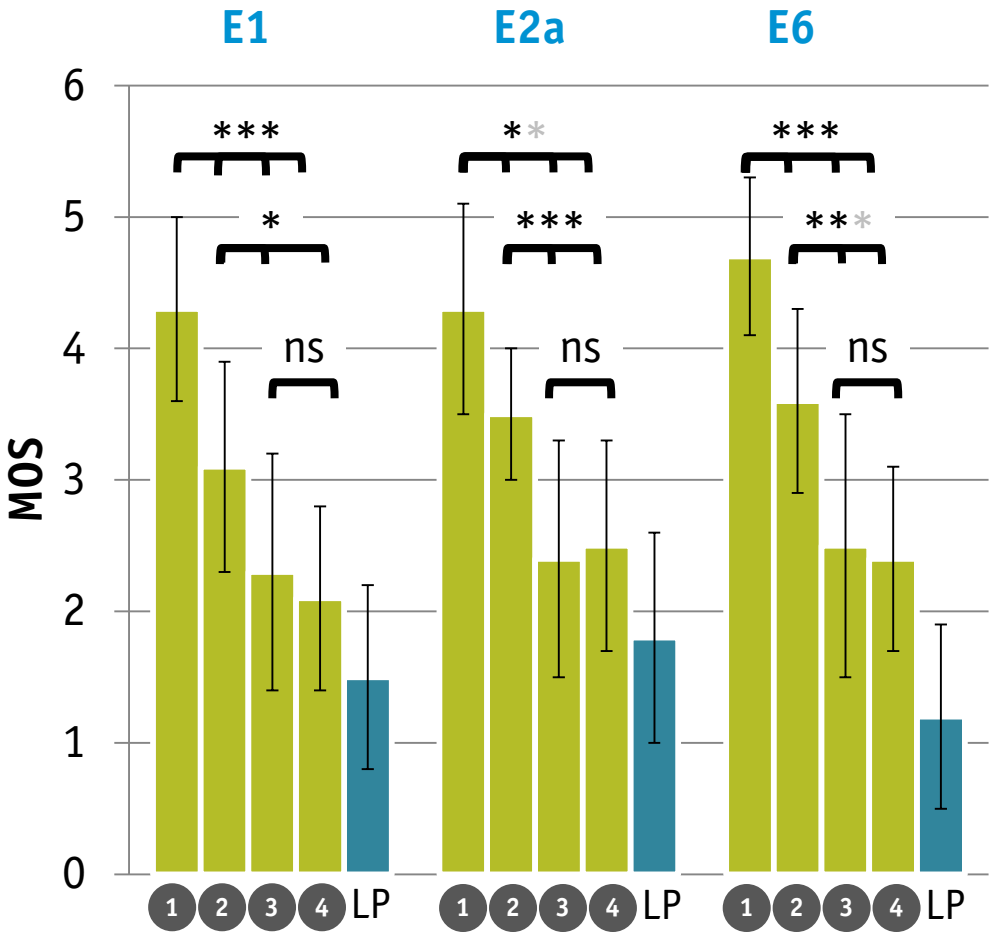
Hypothesis 1: Number of LP Episodes

Hypothesis:

Increasing the number of LP episodes before a **multi-episodic judgment** decreases this judgment.



✔ Hypothesis **partly accepted**

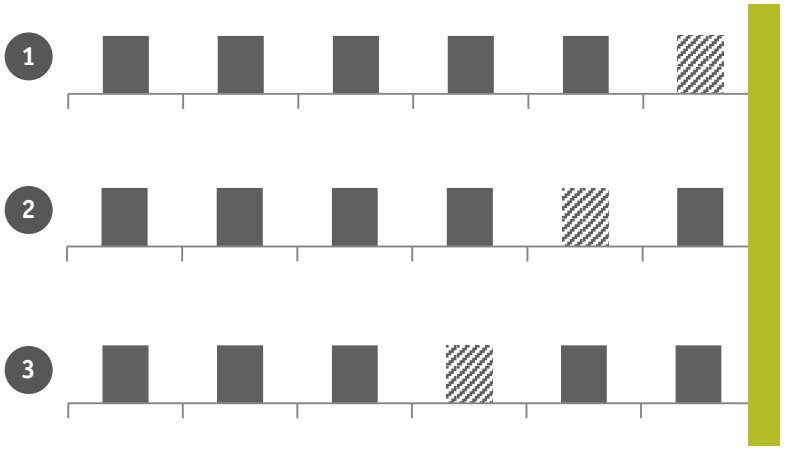


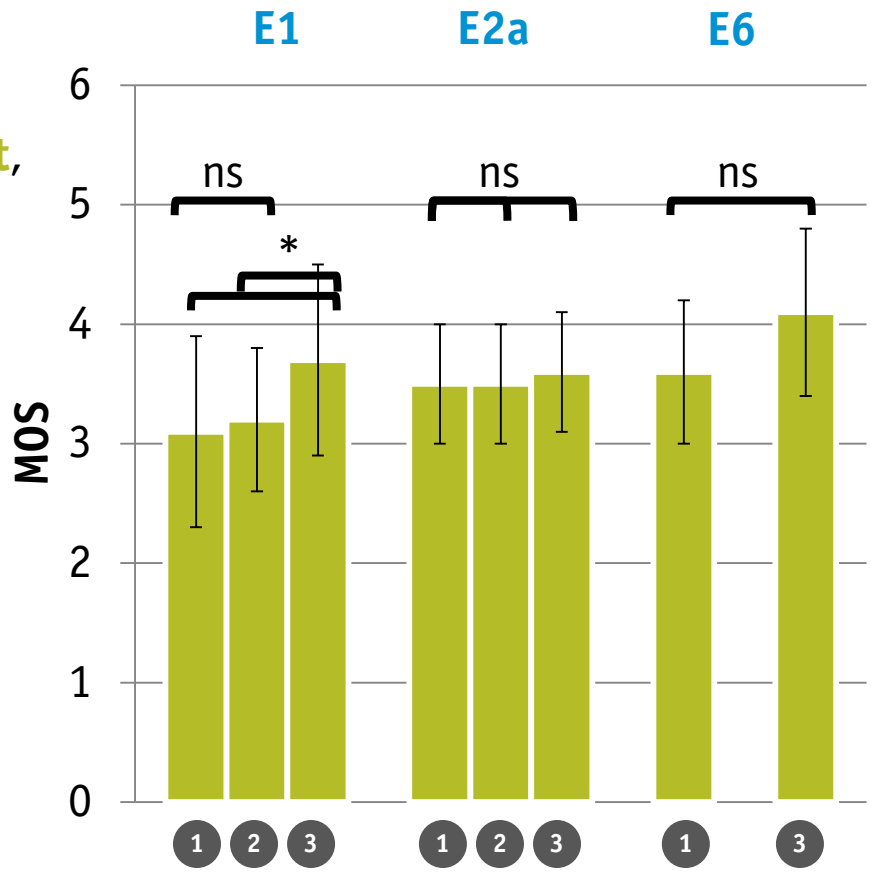
Hypothesis 2: Position of LP Episodes (1)

Hypothesis:

The more HP episodes are presented directly before a **multi-episodic judgment**, the lower is the negative impact of earlier presented LP episodes.

for **one** LP episode





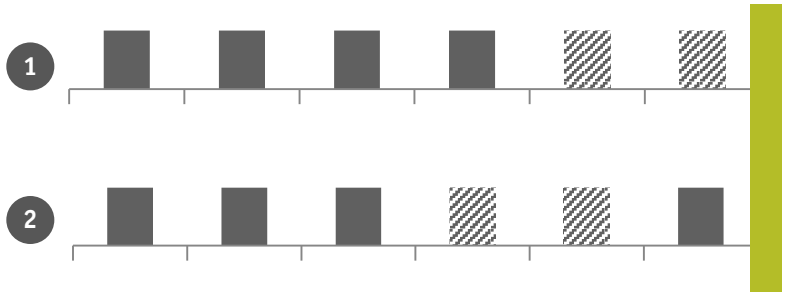
✔ Hypothesis **partly accepted**

Hypothesis 2: Position of LP Episodes (2)

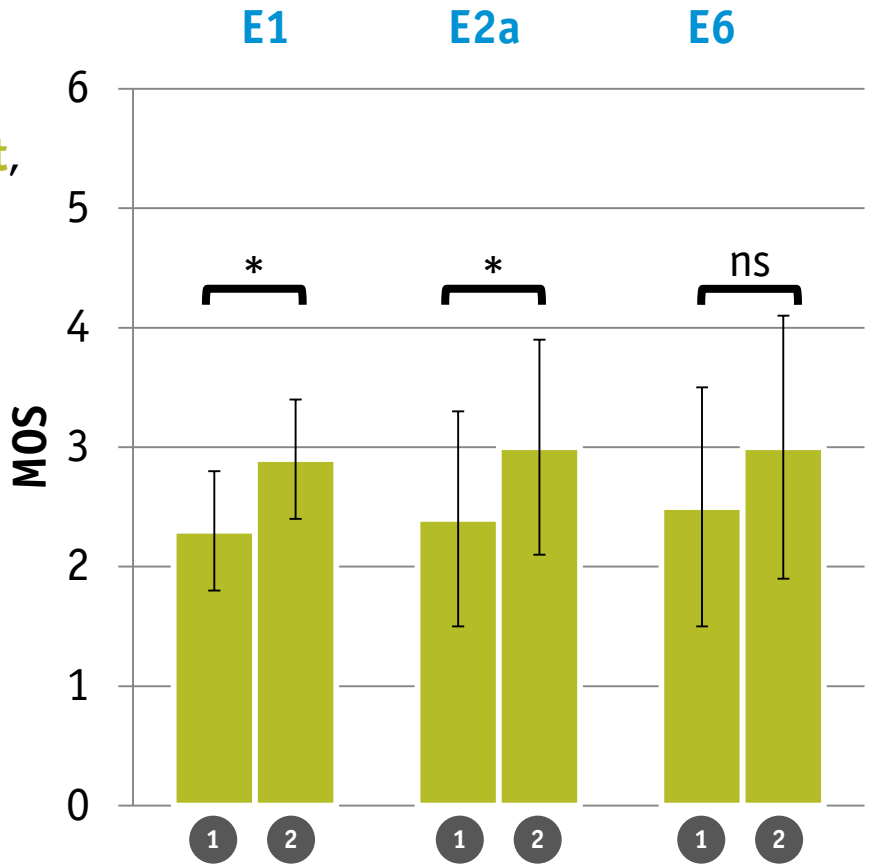
Hypothesis:

The more HP episodes are presented directly before a **multi-episodic judgment**, the lower is the negative impact of earlier presented LP episodes.

for **two** LP episode



✔ Hypothesis **partly accepted**

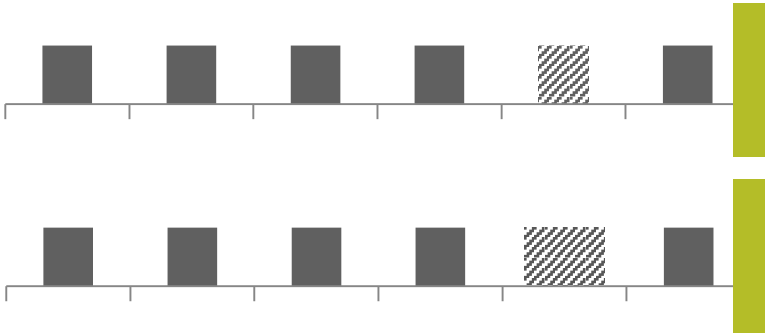


Hypothesis 6: Duration of LP Episodes

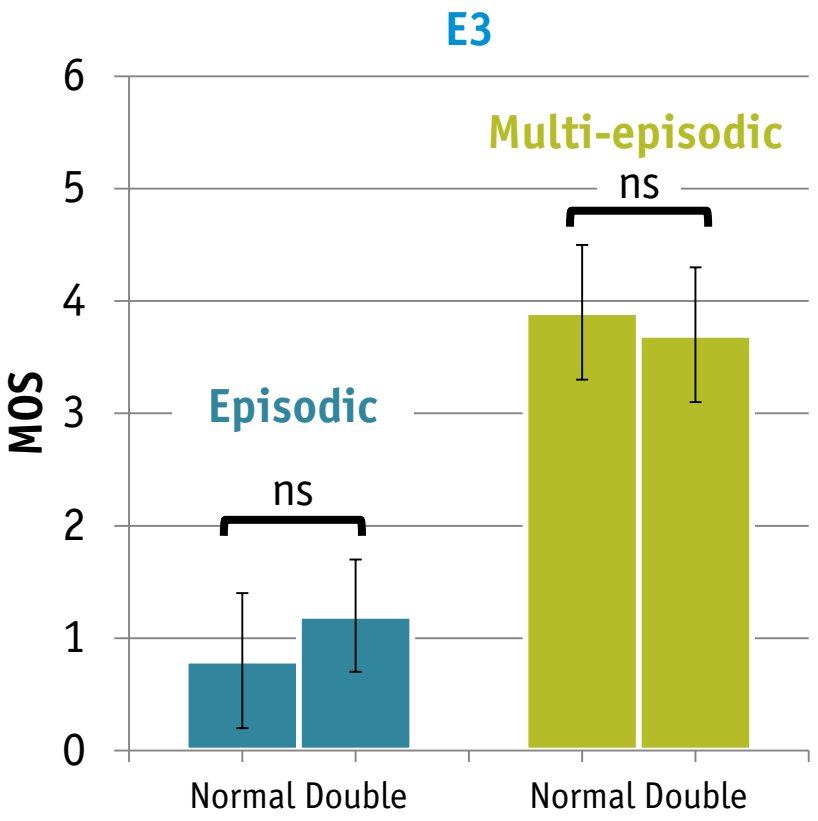
Hypothesis:

LP episodes with a much longer duration result in a higher reduction of **multi-episodic judgments** than shorter LP episodes.

doubled duration of one LP episode



✖ Hypothesis **rejected**



Prediction of Multi-episodic MOS

Goal:

Predict **multi-episodic MOS ($\hat{m}_n$)** based upon **episodic MOS (e_i)** of all prior usage episodes.

Observed effects:

- duration neglect
- position of LP episodes
- number of LP episodes
- saturation

E1, E2a, and E6

Approach

Weighted average

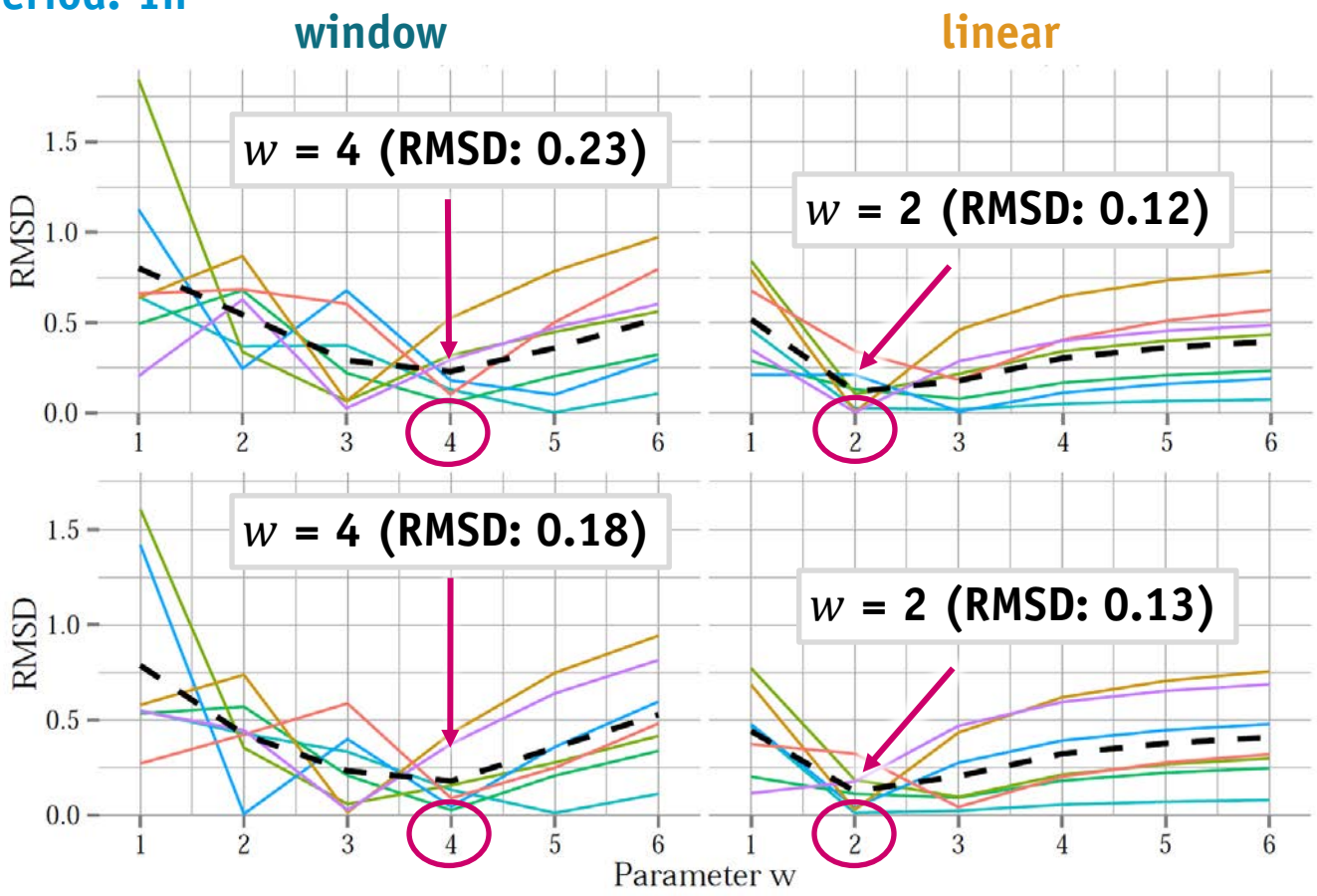
$$\hat{m}_n = \frac{\sum_{i=1}^{n} a_i * e_i}{\sum_{i=1}^{n} a_i}$$

Weight functions: a_i

- **Parameter w:** $1 \leq w \leq n$
- **window**
  linear

Prediction: m_6 (E1 and E2a)

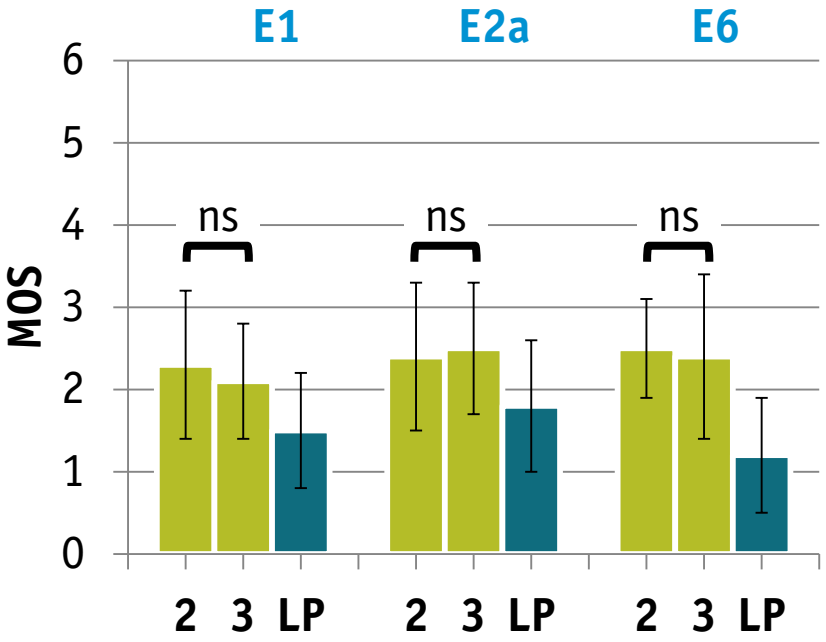
Usage period: 1h



Prediction: Saturation

Saturation observed:

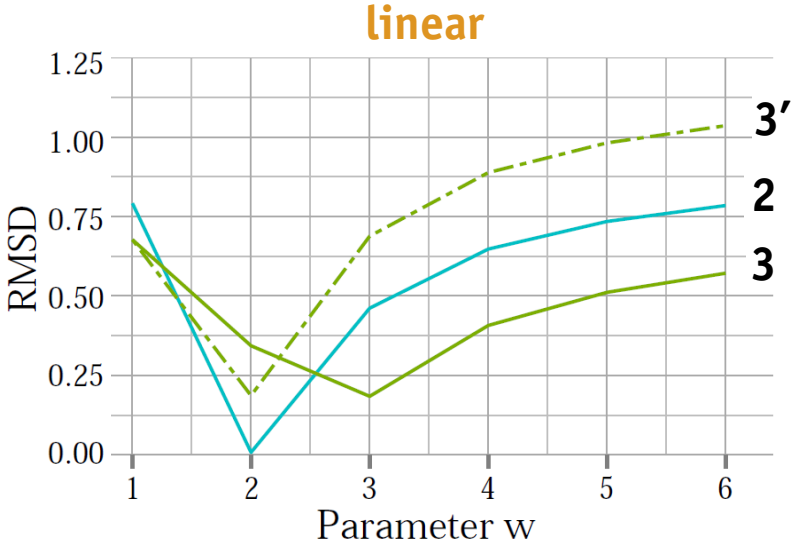
Two and three LP episodes result in a very similar **multi-episodic MOS**



Solution:

Modify **3** with:

$$\textbf{3'}: e_4 := \overline{e_{HP}}$$



Summary and Discussion

#1 Formation process of multi-episodic judgments

- service initially well-working
- severe degradations
- usage periods: 1h and 6 days
- speech-only services

- impact of number
  - **saturation**
- impact of position
- duration neglect

#2 Develop simple, sufficient prediction models

- weighted average
- compensation for saturation

Limitations

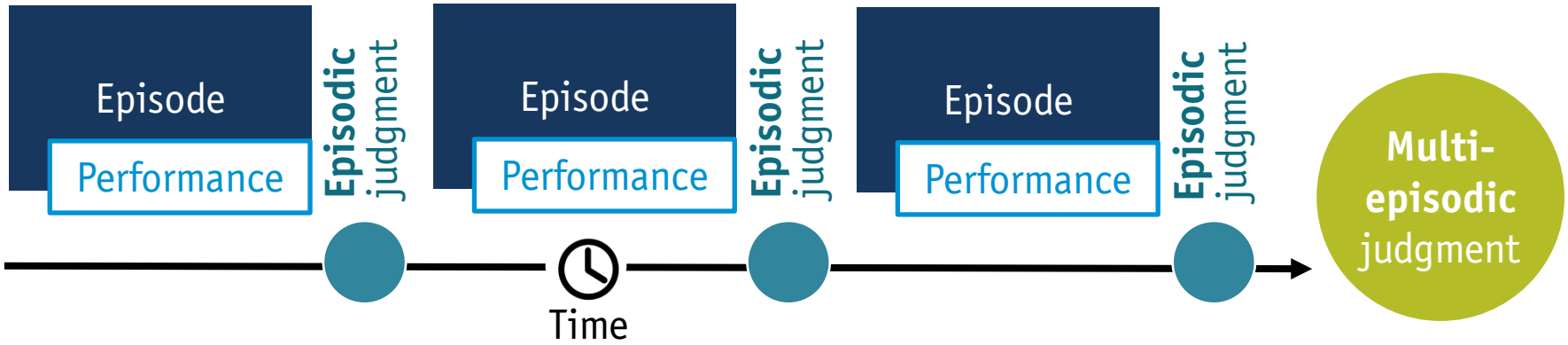
- MOS: *average person*
- sampling of participants
- unrealistic performance

Future Work

Future Work

- non-regular use
- impact of episodic judgments
- within episode performance changes
- importance tasks
- verify findings without defined-use method

Multi-episodic Perceived Quality



References

Möller, Sebastian, Bang, Chihuy, Tamme, Teele, Vaalgamaa, Markus, and Weiss, Benjamin (2011). “From Single-all to Multi-Call Quality: A Study on Long-term Quality Integration in Audio-Visual Speech Communication.” In: 12th Annual Conference of the International Speech Communication Association. INTERSPEECH. Florence, Italy: ISCA, pp. 1485–1488.

Raake, Alexander and Egger, Sebastian (2014). "Quality and Quality of Experience." In: Quality of Experience. Ed. by Sebastian Möller and Alexander Raake. Springer International Publishing, pp. 11–33. isbn: 978-3-319-02681-7

Reiter, Ulrich; Brunnström, Kjell; De Moor, Katrien; Mohamed-Chaker, Larabi; Pereira, Manuela, Pinheiro, Antonio; You, Jungyong and Zgank, Andrej (2014). "Factors Influencing Quality of Experience." In: Quality of Experience. Ed. by Sebastian Möller and Alexander Raake. Springer International Publishing, pp. 55–72. isbn: 978-3-319-02681-7

My Publications

Guse, Dennis and Möller, Sebastian (2013). "Macro-temporal Development of QoE: Impact of Varying Performance on QoE over Multiple Interactions." In: Proceedings of AIA-DAGA Conference on Acoustics. Vol. 46. Merano, Italy: Deutsche Gesellschaft für Akustik, pp. 452–455.

Guse, Dennis, Weiss, Benjamin, and Möller, Sebastian (2014). "Modelling multi-episodic quality perception for different telecommunication services: First insights." In: Sixth International Workshop on Quality of Multimedia Experience (QoMEX). Singapore. IEEE, pp. 105–110.

Weiss, Benjamin; Guse, Dennis; Möller, Sebastian; Raake, Alexander; Borowiak, Adam and Reiter, Ulrich (2014). "Temporal Development of Quality of Experience." In: Quality of Experience. SpringerInternational Publishing, pp. 133–147. isbn: 978-3-319-02681-7.

Guse, Dennis; Wunderlich, Anna; Weiss, Benjamin; Möller Sebastian (2016). "Duration Neglect in Multi-episodic Perceived Quality". In: *Proc. International Conference on Quality of Multimedia Experience (QoMEX)*.

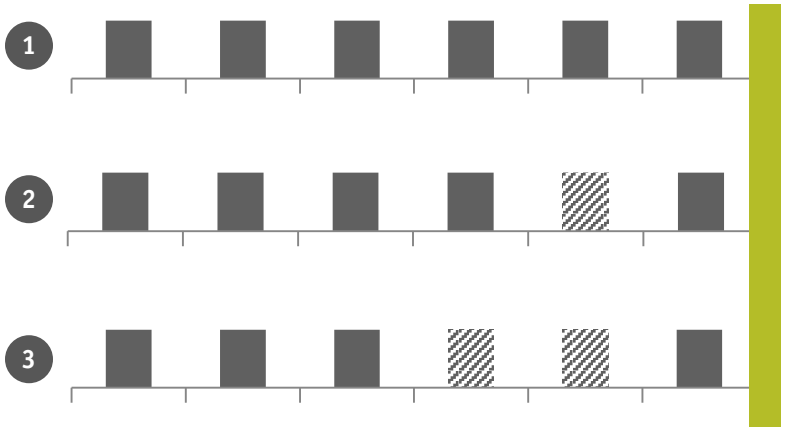
Performance Levels

Performance Level	Signal Bandwidth	Codec	POLQA Estimation
High Performance (HP)	50..700 Hz	G.722, Mode 1	4.0
Medium Performance (MP)	300..3400 Hz	G.711	3.3
Low Performance (LP)	300..3400 Hz	LPC-10	1.9

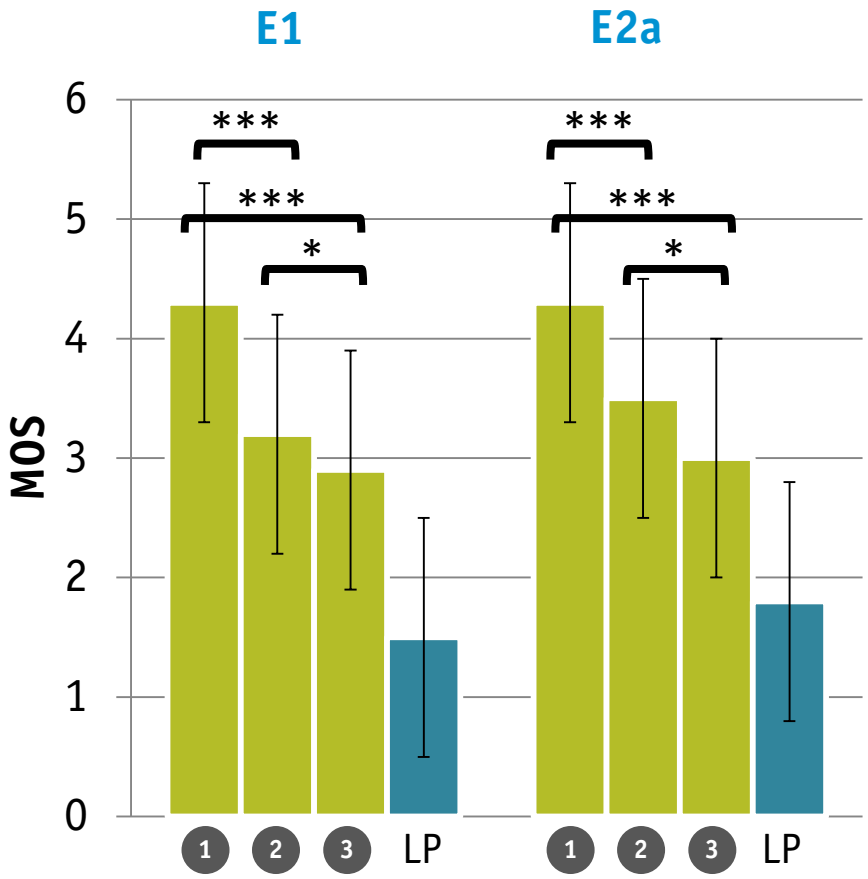
Hypothesis 1: Number of LP Episodes (1)

Hypothesis:

Increasing the number of LP episodes before a multi-episodic judgment decreases this judgment.



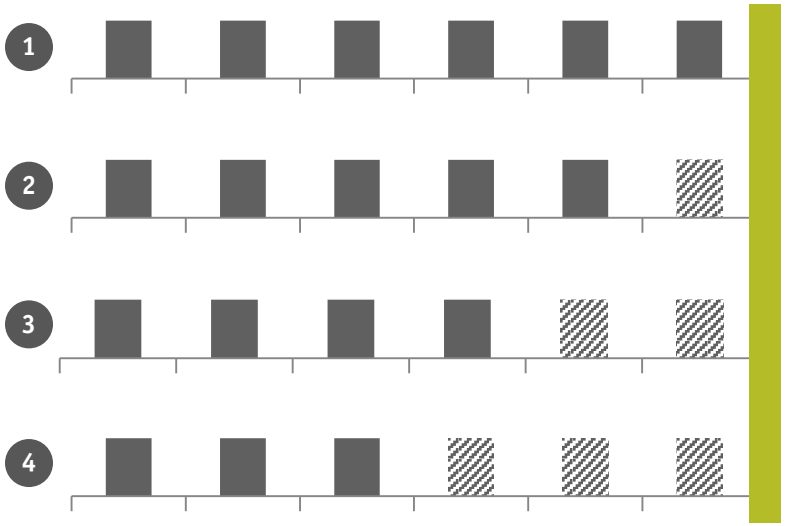
✔ Hypothesis **accepted**



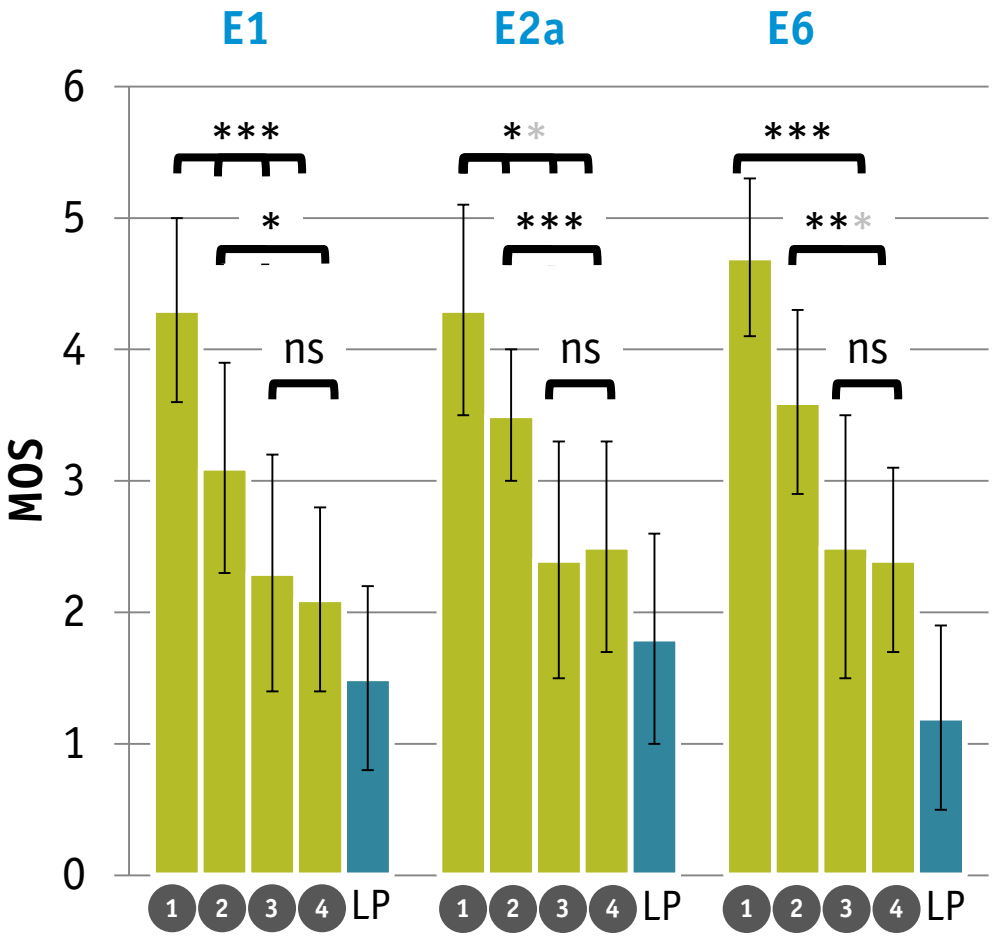
Hypothesis 1: Number of LP Episodes (2)

Hypothesis:

Increasing the number of LP episodes before a multi-episodic judgment decreases this judgment.



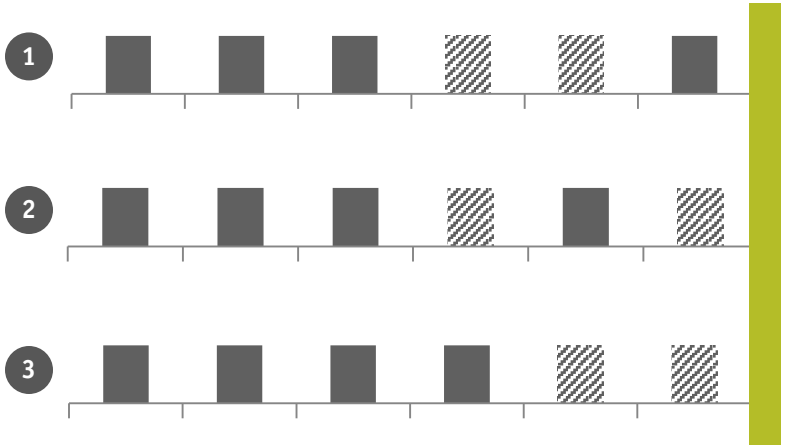
✔ Hypothesis **partly accepted**



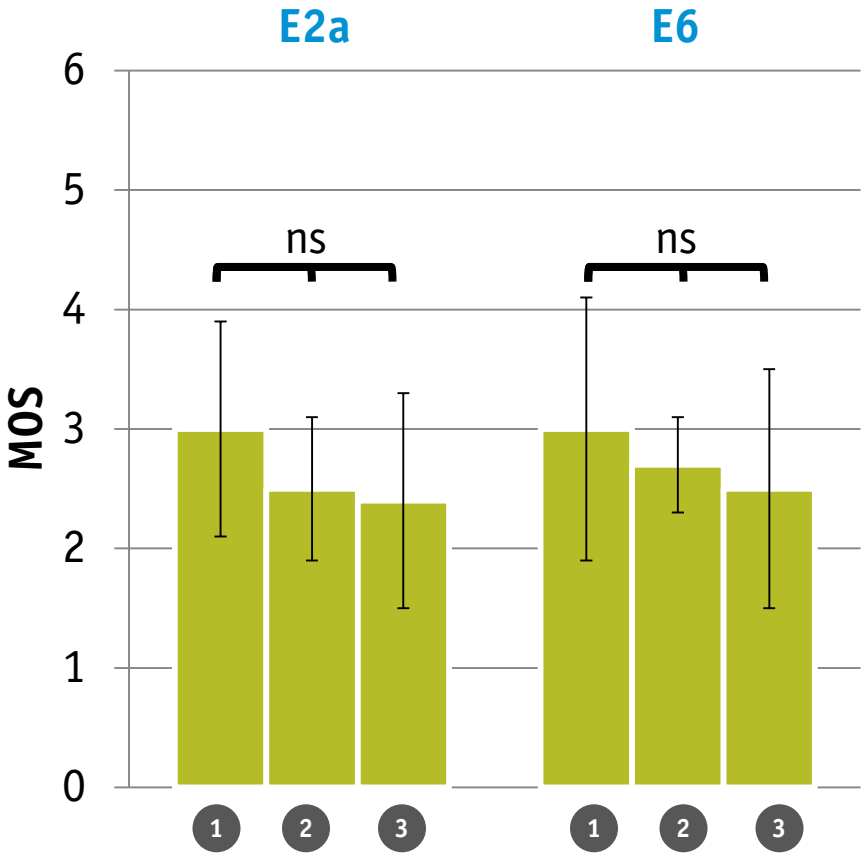
Hypothesis 3: (Non)-Consecutive LP Episodes

Hypothesis:

The presentation of non-consecutive LP episodes leads to a higher reduction of multi-episodic judgments than a consecutive presentation.

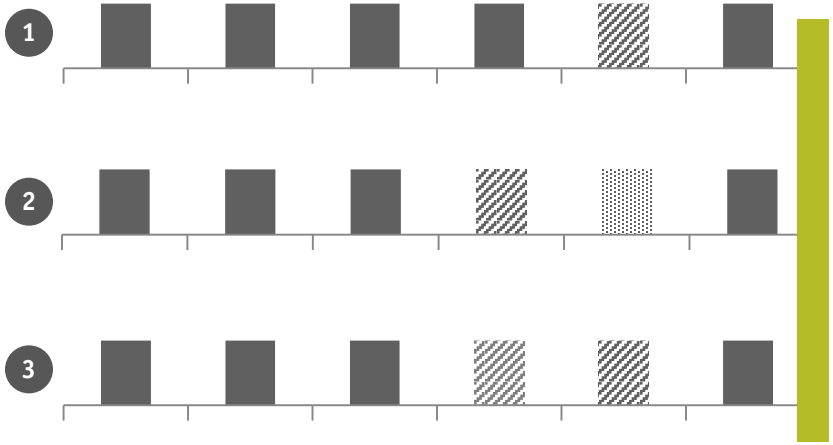


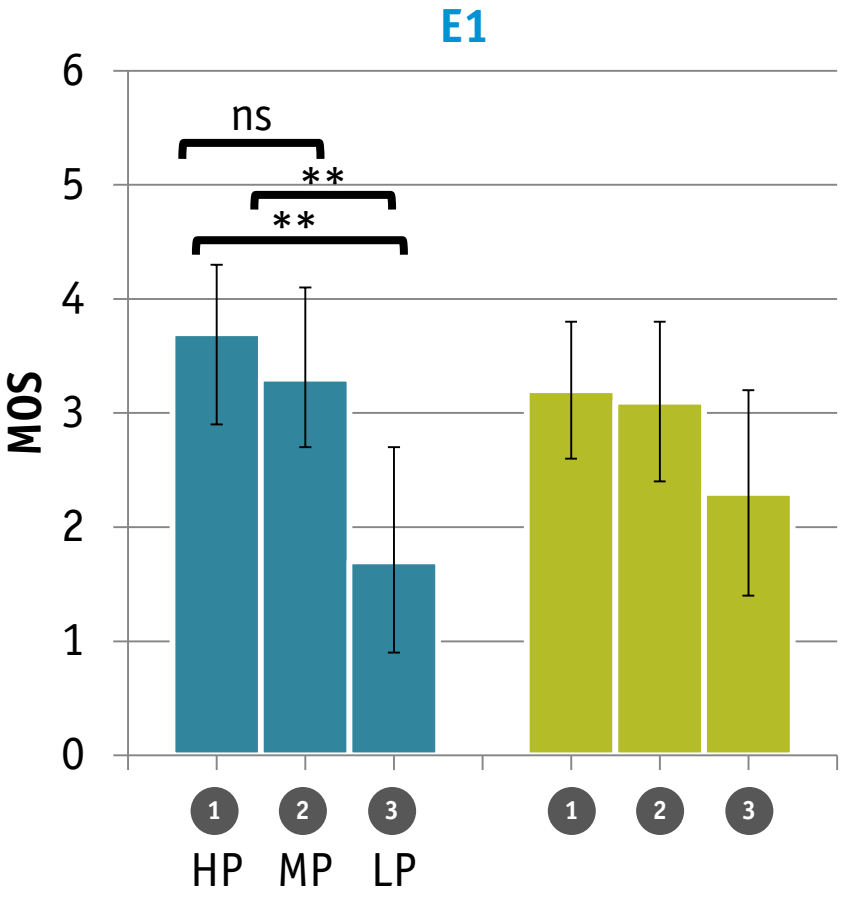
✖ Hypothesis **rejected**



Hypothesis 4: Strength of Degradation

Hypothesis:
The lowest experienced episodic performance has an increased impact on multi-episodic judgments, whereas less severe degradations are less important.

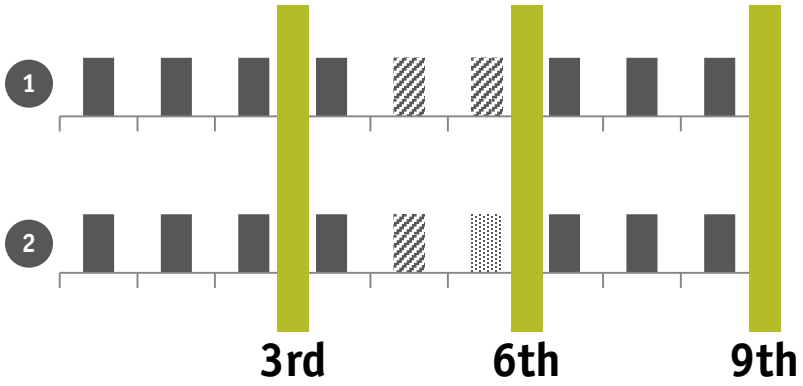




Hypothesis 5: Recovery

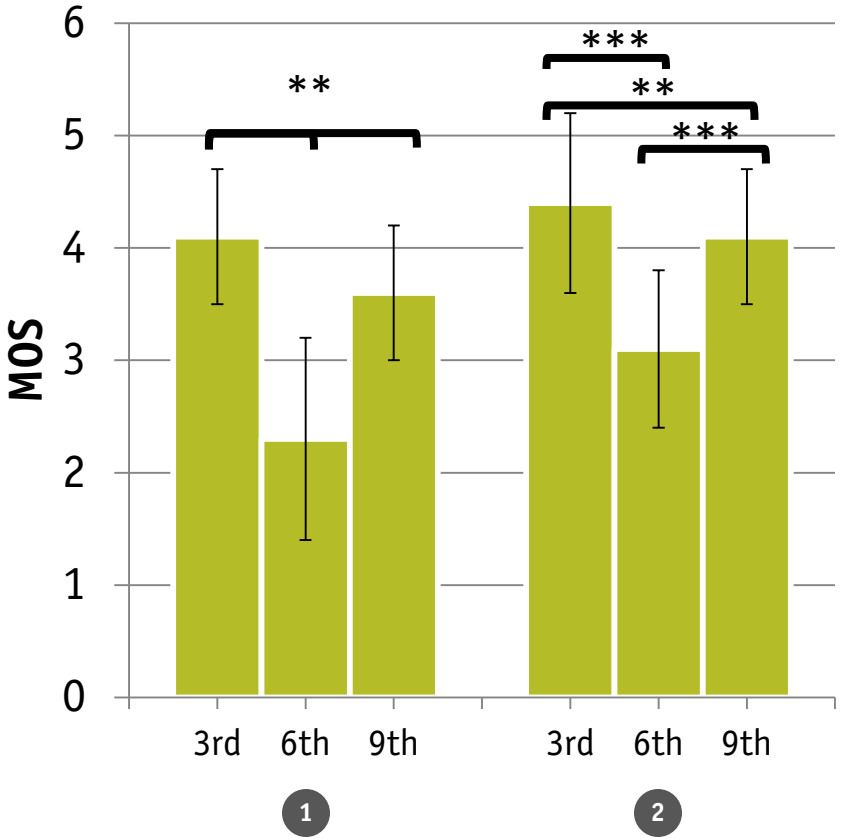
Hypothesis:

Presenting additional HP episodes after a negatively affected multi-episodic judgments results in an increase of the following multi-episodic judgment.



✔ Hypothesis **accepted**

E1

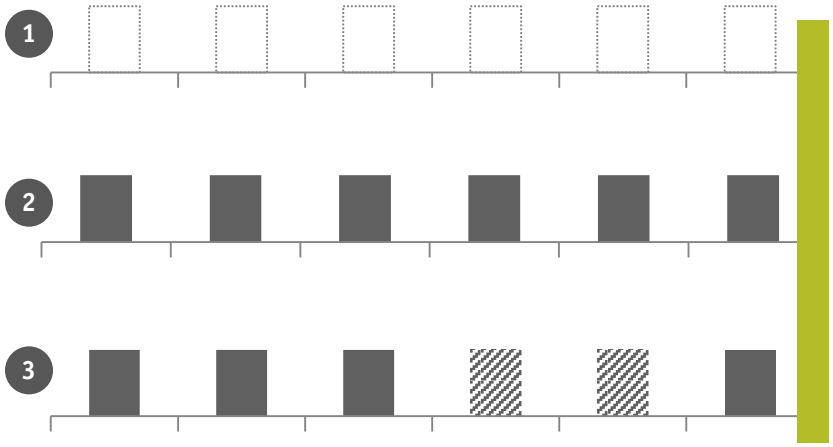


Hypothesis 7: Impact of a 2nd Service

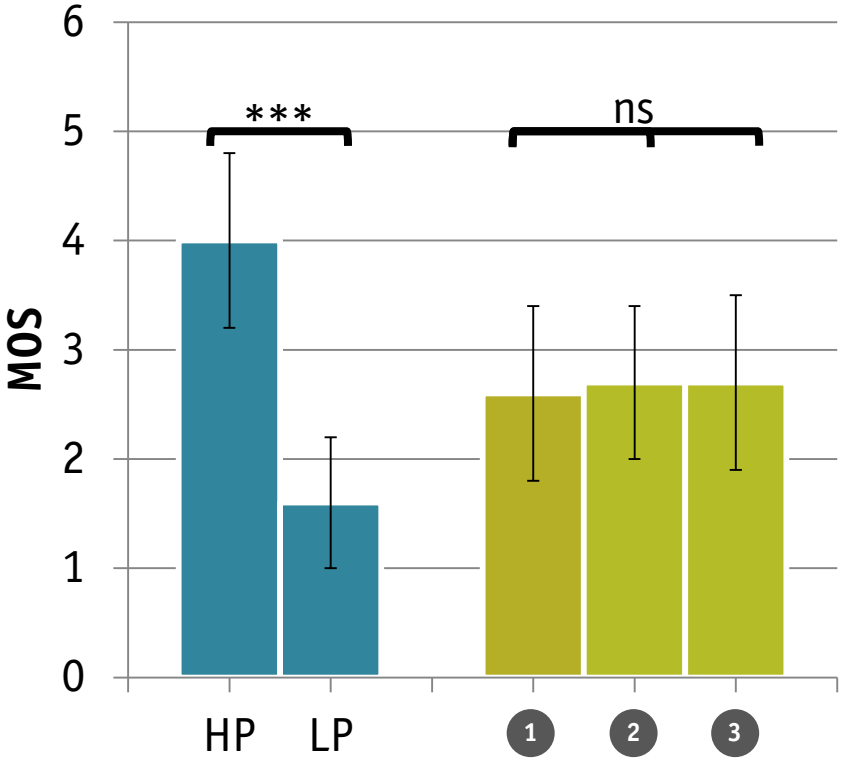
Hypothesis:

The multi-episodic judgment for one service is not affected by the presentation of a 2nd service in the same usage period.

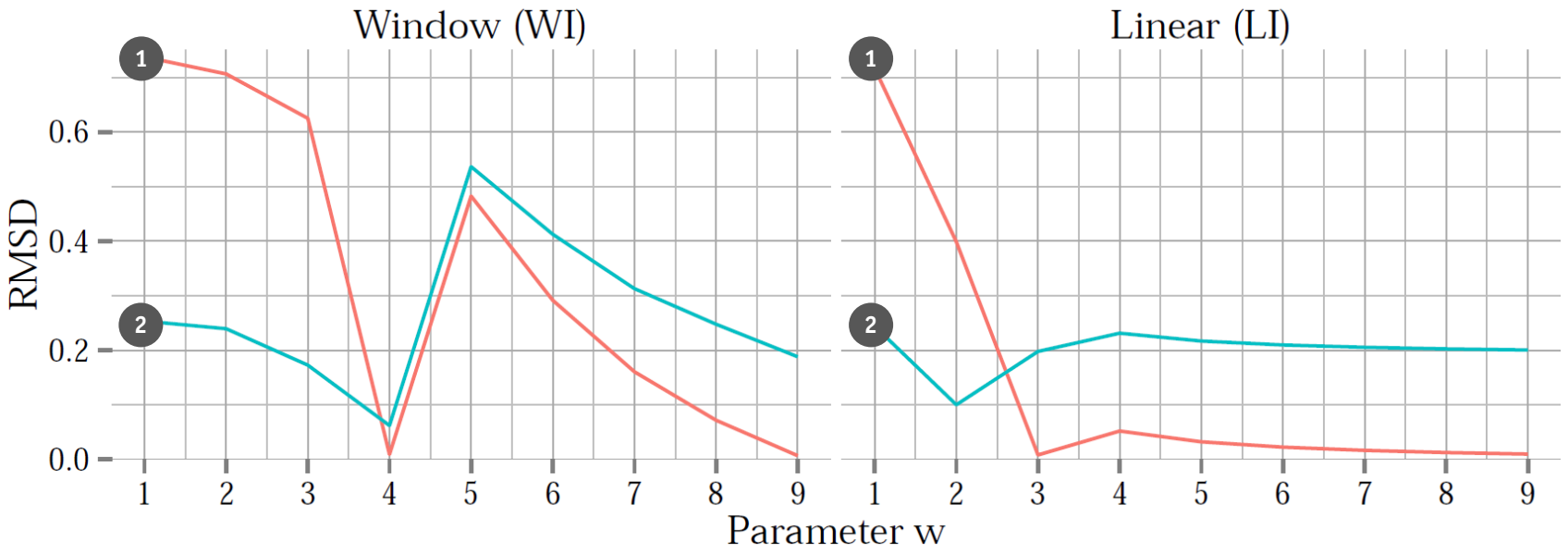
Sequential use of two services



E2b

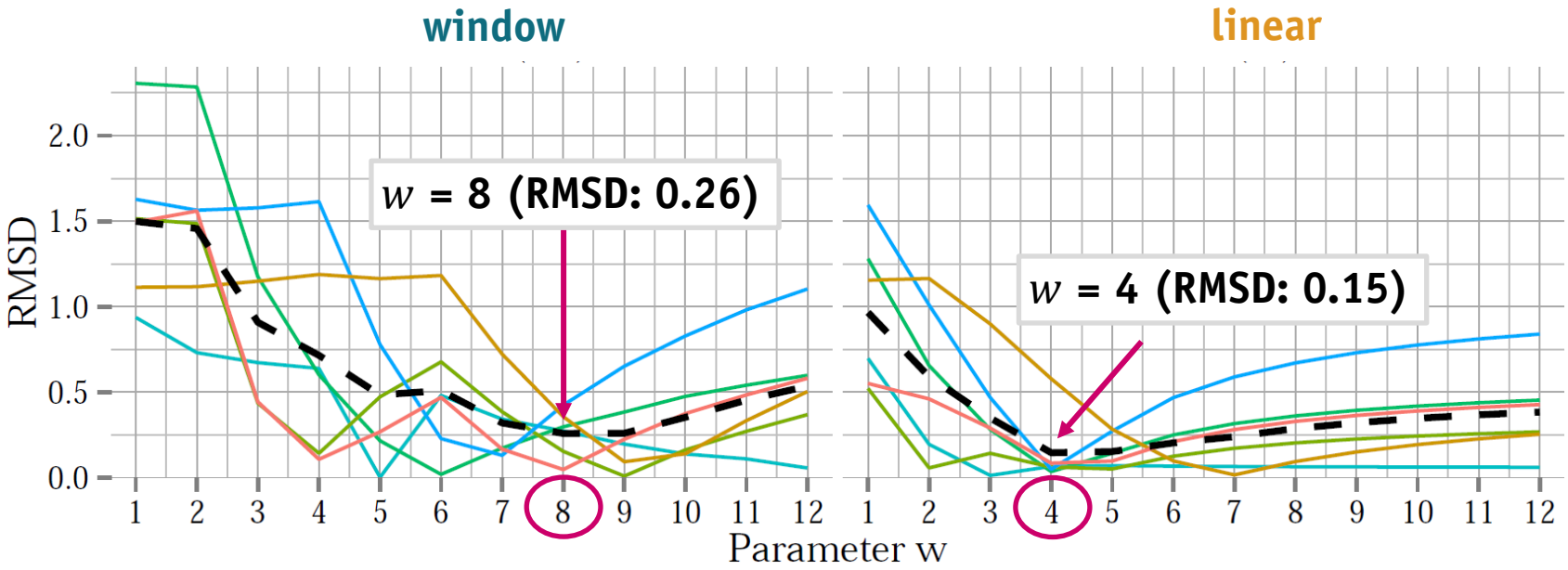


Prediction: m_9 (E1)



Prediction: m_{12} (E6)

Usage Period: 6 days (E6)



Quality of Experience (Definition)

"Quality of Experience is the degree of delight or annoyance of a person whose experiencing involves an application, service, or system.

It results from the person's evaluation of the fulfillment of his or her expectations and needs with respect to the utility and/or enjoyment in the light of the person's context, personality and current state."

(Raake et al., 2014, p. 19)