Perceiving fine-phonetic detail: The case of word-final /s/ in English

Dr Dominic Schmitz
English Language and Linguistics
Heinrich Heine University Düsseldorf

dominic.schmitz@hhu.de
dmncschmtz.bsky.social

Why subphonemic detail matters

- speech contains rich, fine-grained phonetic detail
- classic models assume early abstraction to phonemes
- subphonemic variation is often treated as noise
- but production data tell a different story
- where do we expect the boundary between gradient detail and noise?

e.g. Chomsky and Halle (1968); Kiparsky (1982); Levelt et al. (1999); Roelofs and Ferreira (2019); Turk and Shattuck-Hufnagel (2020)

Why subphonemic detail matters

- speech contains rich, fine-grained phonetic detail
- classic models assume early abstraction to phonemes
- subphonemic variation is often treated as noise
- but production data tell a different story
- where do we expect the boundary between gradient detail and noise?

e.g. Chomsky and Halle (1968); Kiparsky (1982); Levelt et al. (1999); Roelofs and Ferreira (2019); Turk and Shattuck-Hufnagel (2020)

Why subphonemic detail matters

- speech contains rich, fine-grained phonetic detail
- classic models assume early abstraction to phonemes
- subphonemic variation is often treated as noise
- but production data tell a different story
- where do we expect the boundary between gradient detail and noise?

e.g. Chomsky and Halle (1968); Kiparsky (1982); Levelt et al. (1999); Roelofs and Ferreira (2019); Turk and Shattuck-Hufnagel (2020)

Why subphonemic detail matters

- speech contains rich, fine-grained phonetic detail
- classic models assume early abstraction to phonemes
- subphonemic variation is often treated as noise
- but production data tell a different story
- where do we expect the boundary between gradient detail and noise?

e.g. Chomsky and Halle (1968); Kiparsky (1982); Levelt et al. (1999); Roelofs and Ferreira (2019); Turk and Shattuck-Hufnagel (2020)

Subphonemic regularities in production

- across multiple domains, speakers produce
  - subtle probabilistic lengthening
  - systematic differences tied to morphological category
  - cross-linguistic consistency of patterns
- examples
  - English word-final /s/
  - German word-final /s/
  - German a-schwa in masculine nouns
- these patterns are too **systematic** to be dismissed as random phonetic noise

Schmitz et al. (2021); Schmitz & Bear-Henney (2024); Schmitz (in prep.)

Do listeners perceive these subtle differences?

- production clearly encodes structure
- but
  - are these cues perceptually available?
  - if yes, under what conditions?
  - is there a threshold for perception?
  - how does this challenge models of speech perception?

Roadmap

1. production evidence across English and German
2. why this challenges abstractionist models
3. perception experiment on English word-final /s/
    a) experiment
    b) results
4. integrating production, perception, and representation
5. cross-linguistic and theoretical implications

Production evidence for subphonemic morphology effects

- speakers do not pronounce identical segment strings in identical ways
- small, systematic duration differences appear depending on morphological category

  = subphonemic differences: smaller than a phoneme, not contrastive in the phonological system, but still physically real
- these differences are measurable across corpora and lab experiments
- they occur even when speakers are unaware of producing any difference

Plag et al. (2017); Zimmermann et al. (2016); Schmitz et al. (2021); Schmitz & Bear-Henney (2024); Schmitz (in prep.)

English /s/ duration hierarchy

- English words can end in /s/ that is
  - non-morphemic (part of the stem: *bus*, *grass*, *lens*)
  - plural (*books*, *cats*)
  - clitic (*John's*, *that's*)
- all have the same phonological shape, but differ in function
- production shows a striking pattern

??? /s/ > ??? /s/ > ??? /s/

Plag et al. (2017); Schmitz et al. (2021)

English /s/ duration hierarchy

- English words can end in /s/ that is
  - non-morphemic (part of the stem: *bus*, *grass*, *lens*)
  - plural (*books*, *cats*)
  - clitic (*John's*, *that's*)
- all have the same phonological shape, but differ in function
- production shows a striking pattern

  non-morphemic /s/ > plural /s/ > clitic /s/

- the pattern holds across many datasets, varieties, and methodologies

Plag et al. (2017); Schmitz et al. (2021)

Why should /s/ duration depend on morphology?

- traditional models of speech production assume that
    1. all /s/ sounds should be identical at the phonological level
    2. phonetics should not know about morphology
- yet speakers consistently encode morphological information in duration
- suggests that gradient phonetic detail is meaningfully structured
- challenges the idea that surface phonetics is independent of lexical representations

e.g. Chomsky and Halle (1968); Kiparsky (1982); Levelt et al. (1999); Roelofs and Ferreira (2019); Turk and Shattuck-Hufnagel (2020)

Similar patterns in German reinforce the case

- German also allows both non-morphemic and plural /s/
- production shows

  non-morphemic /s/ > plural /s/

  in duration, mirroring English
- this **cross-linguistic recurrence** suggests a general phenomenon
- when two languages with very different morphological systems show the same pattern, it points to a deeper cognitive mechanism
- duration is systematically shaped by morphological status, not random noise

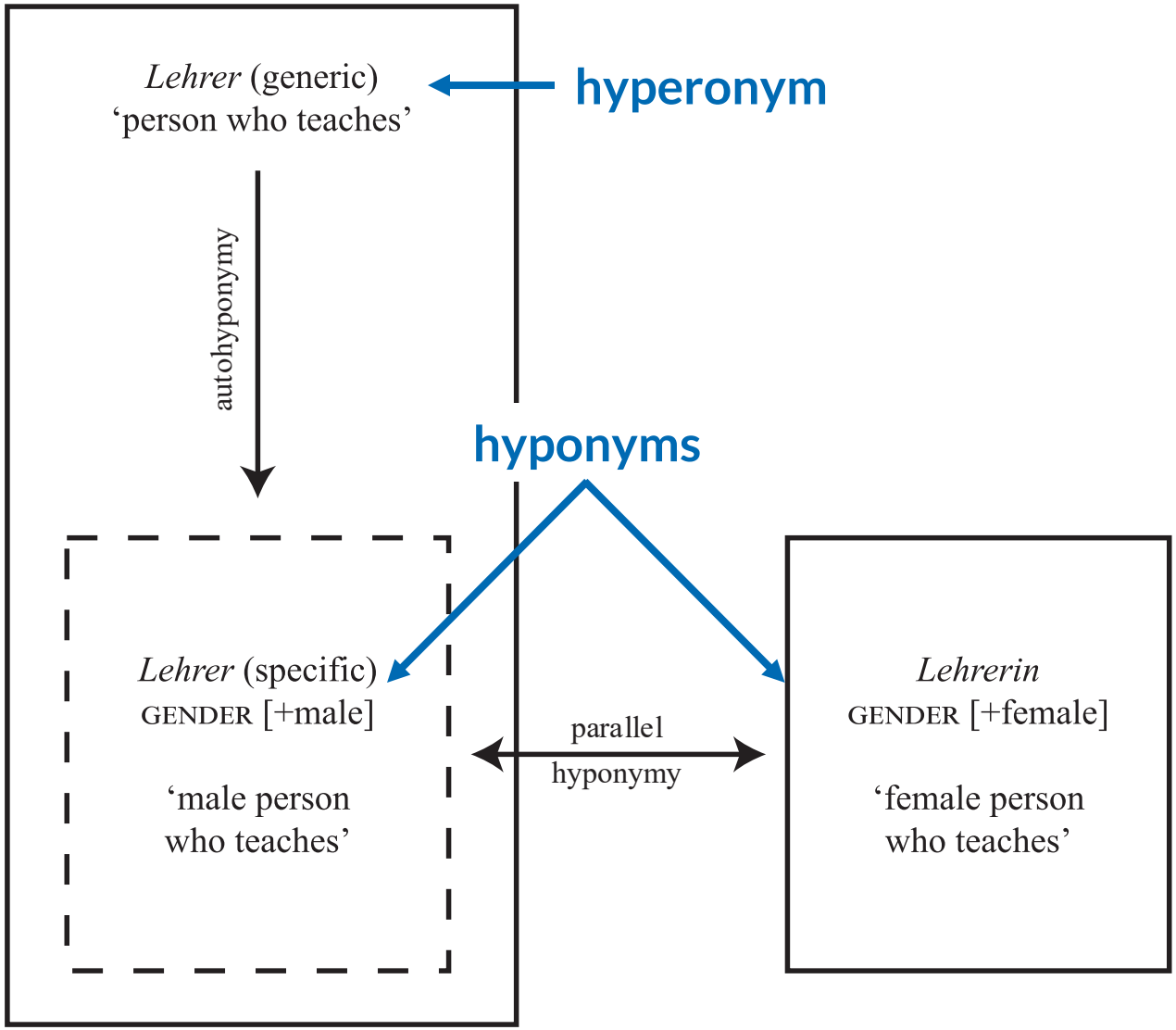
Schmitz & Bear-Henney (2024); Baer-Henney & Schmitz (submitted)

Subphonemic detail beyond consonants: the a-schwa case

- many German masculine nouns are **autohyponyms**
  - generic meaning (*der Lehrer* = ‘person who teaches’)
  - specific meaning (*der Lehrer* = ‘male person who teaches’)

Schmitz (in prep.)

Subphonemic detail beyond consonants: the a-schwa case



Schmitz (in prep.)

Subphonemic detail beyond consonants: the a-schwa case

- many German masculine nouns are **autohyponyms**
  - generic meaning (*der Lehrer* = ‘person who teaches’)
  - specific meaning (*der Lehrer* = ‘male person who teaches’)
- both readings are **segmentally identical**
- but speakers produce longer /ɐ/ in the generic reading
- duration correlates with the semantic structure of the lexicon
- shows that even **sense distinctions can surface as phonetic differences**
- this suggests subphonemic duration is linked not only to morphology, but to semantics

Schmitz (in prep.)

What production evidence tells us so far

- duration differences are systematic and linguistically meaningful
- they appear in multiple languages and morphological systems
- they reflect relationships between form and meaning
- they motivate a key question:

  if speakers encode this detail, can listeners perceive it?

- what do models of speech processing predict?

The abstractionist view in speech perception

- models like TRACE and Shortlist B assume early mapping to phonemes
- this is what is called **early abstraction**: the acoustic signal is quickly reduced to phoneme symbols
- subphonemic variation is treated as noise to be normalised away
- the system focuses on stable phonological categories
- **prediction**: listeners should not (or really, cannot) rely on tiny duration differences for processing

e.g. McClelland & Elman (1986); Norris & McQueen (2008)

What abstractionist models predict about duration cues

- since all /s/ tokens share the same phoneme, duration differences should be ignored
- morphological differences should not affect phonetic encoding
- listeners should not be able to use subphonemic duration for decisions
- duration contrasts (e.g. 10–30 ms) should not influence lexical processing
- so, if listeners *do* perceive some duration contrasts, abstraction **cannot** account for this

e.g. McClelland & Elman (1986); Norris & McQueen (2008)

The emergentist perspective on speech processing

- speech perception reflects learned associations between form and meaning
- fine phonetic detail is part of these associations, not stripped away
- no strict boundary between phonetics and phonology
- patterns like longer non-morphemic /s/ emerge from lexical structure

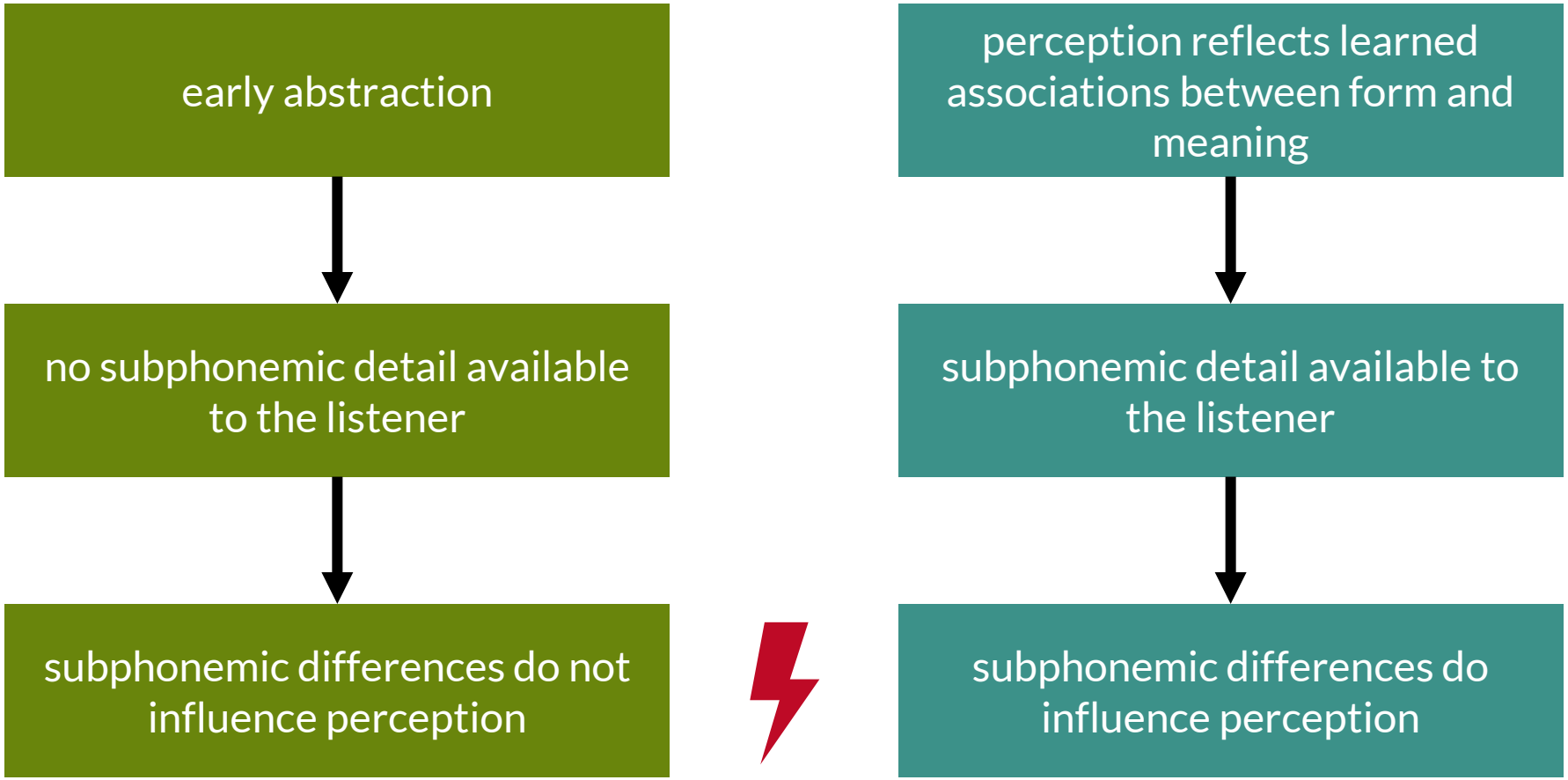
e.g. Baayen et al. (2019); Beach et al. (2021)

What emergentist models predict about duration cues

- listeners may retain and use some subphonemic cues
- perception thresholds may limit which cues are accessible
- the perceptual system might track gradient differences related to morphology
- lexical organisation can shape both production and perception
- so, if listeners *do* perceive some duration contrasts, emergentist models **can** account for this

e.g. Baayen et al. (2019); Beach et al. (2021)

Abstractionist vs emergentist view



so, who's right?

A perception experiment

- production clearly encodes morphological and semantic distinctions
- theory tells us listeners might or might not access these cues
- empirical data are required to determine
  - which duration differences listeners can hear
  - how large the difference must be

Basic idea of the experiment

- uses a well-established perceptual paradigm: the **same-different task**
- listeners hear two stimuli in sequence
- their task is to decide whether the stimuli sound identical or different
- this method measures perceptual sensitivity to acoustic differences
- it is widely used in phonetics and psycholinguistics

Schmitz (2022)

Real words and pseudowords as stimuli

- the perception study targets word-final /s/ duration in…
  - **real words**: allows us to investigate /s/ duration in familiar morphological contexts
  - **pseudowords**: removes lexical, semantic, and frequency effects
- combining both lets us test whether perception depends on lexical knowledge or purely acoustic cues

Schmitz (2022)

Stimuli: real words and pseudowords

non-morphemic /s/	
box	hoax
coax	mix
corpse	tax

plural /s/	
steps	parts
points	rights
groups	books

pseudowords					
bloups	glaips	pleeps	glips	cloops	prups
blouts	glaits	pleets	glits	cloots	pruts
blouks	glaiks	pleeks	gliks	clooks	pruks
bloufs	glaifs	pleefs	glifs	cloofs	prufs

Schmitz (2022)

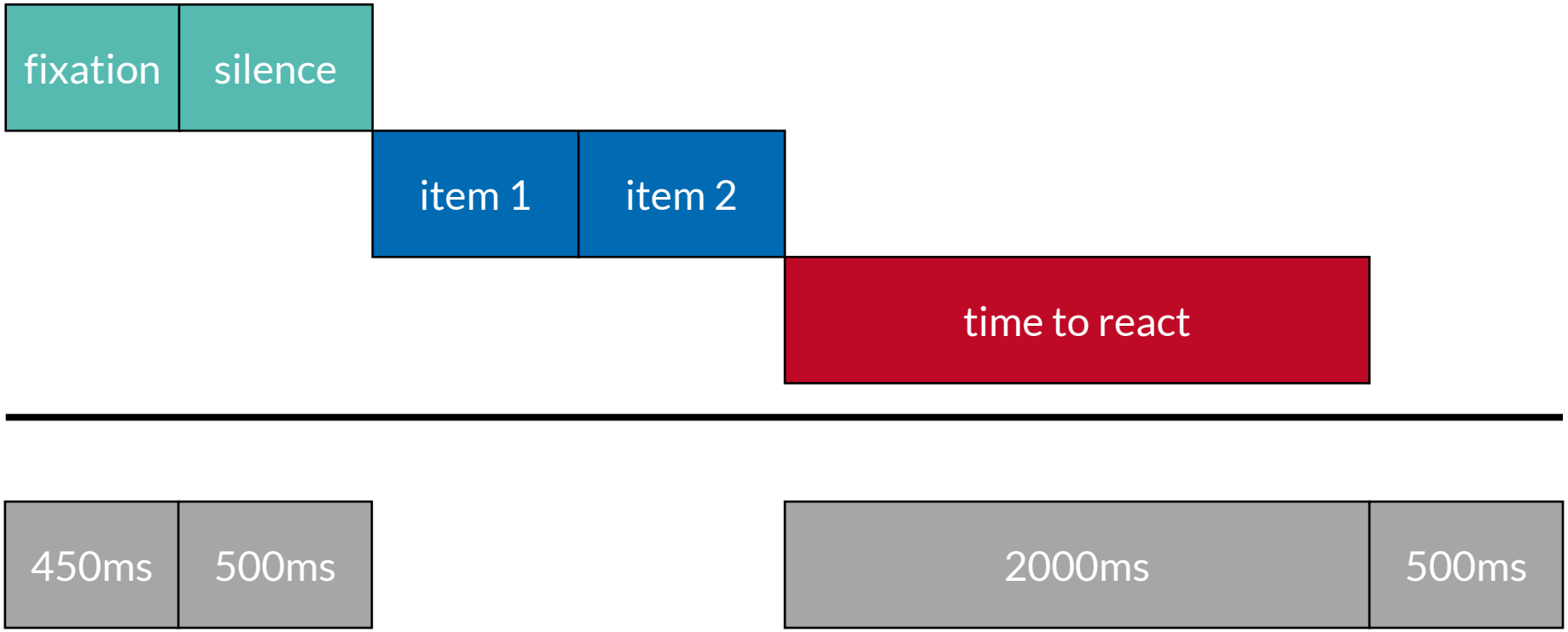
Durational manipulation of /s/

- A prototypical length
- B nm -10; pl +10
- C nm -20; pl +20
- D nm -35; pl +35
- E nm -75; pl +75

pair	same/different	durational dif
A+B	different	±10ms
A+C	different	±20ms
A+D	different	±35ms
A+E	different	±75ms
A+A	same	none
B+B	same	none
C+C	same	none
D+D	same	none
E+E	same	none

Schmitz (2022)

Trial structure



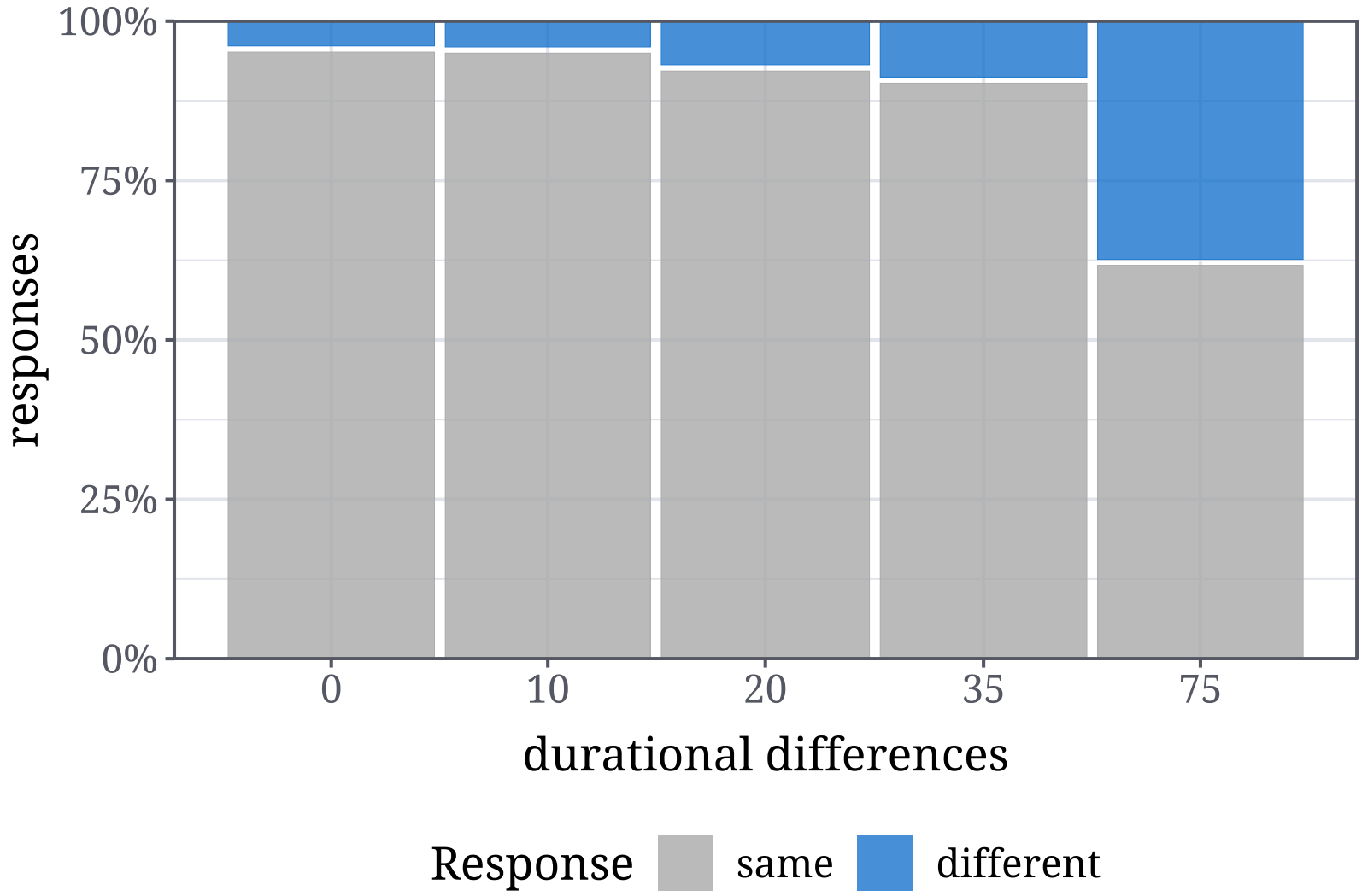
Schmitz (2022)

What participants saw

same	different
A	K

Schmitz (2022)

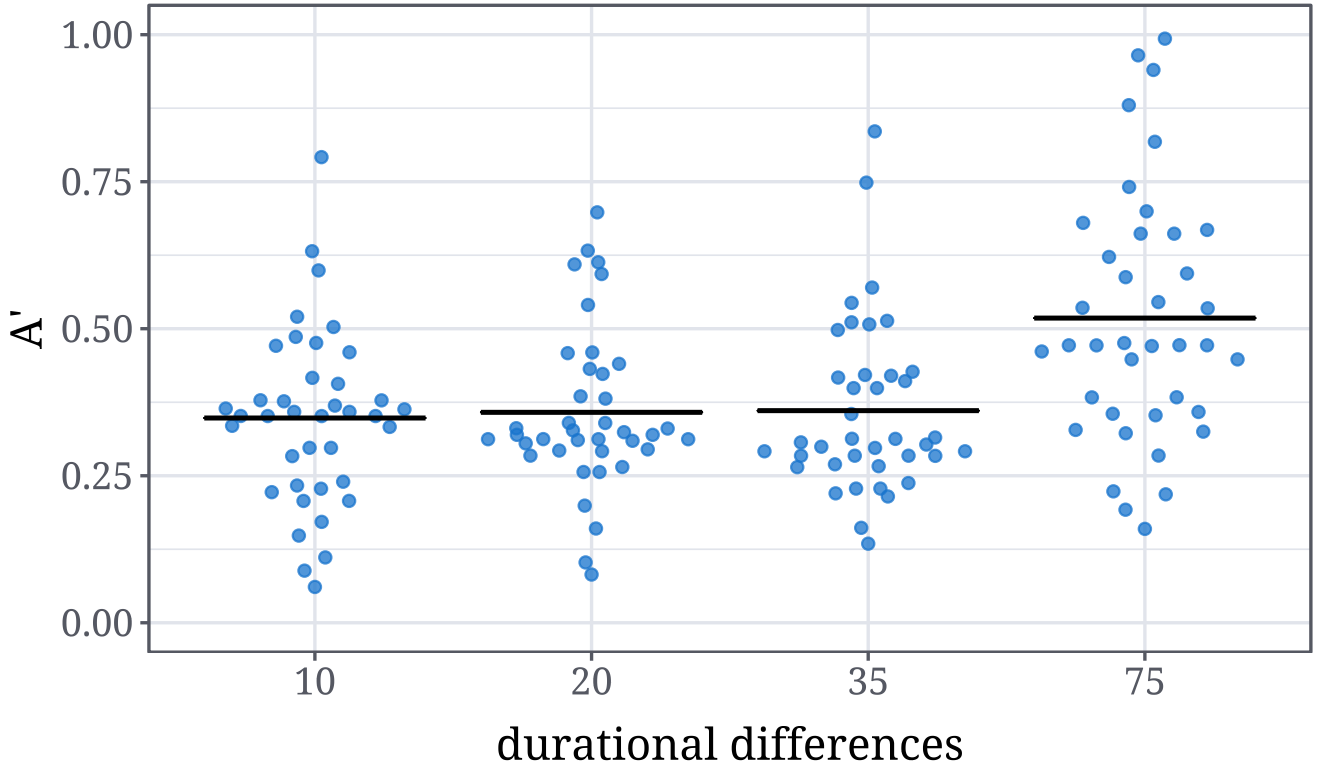
Overview of responses



Schmitz (2022)

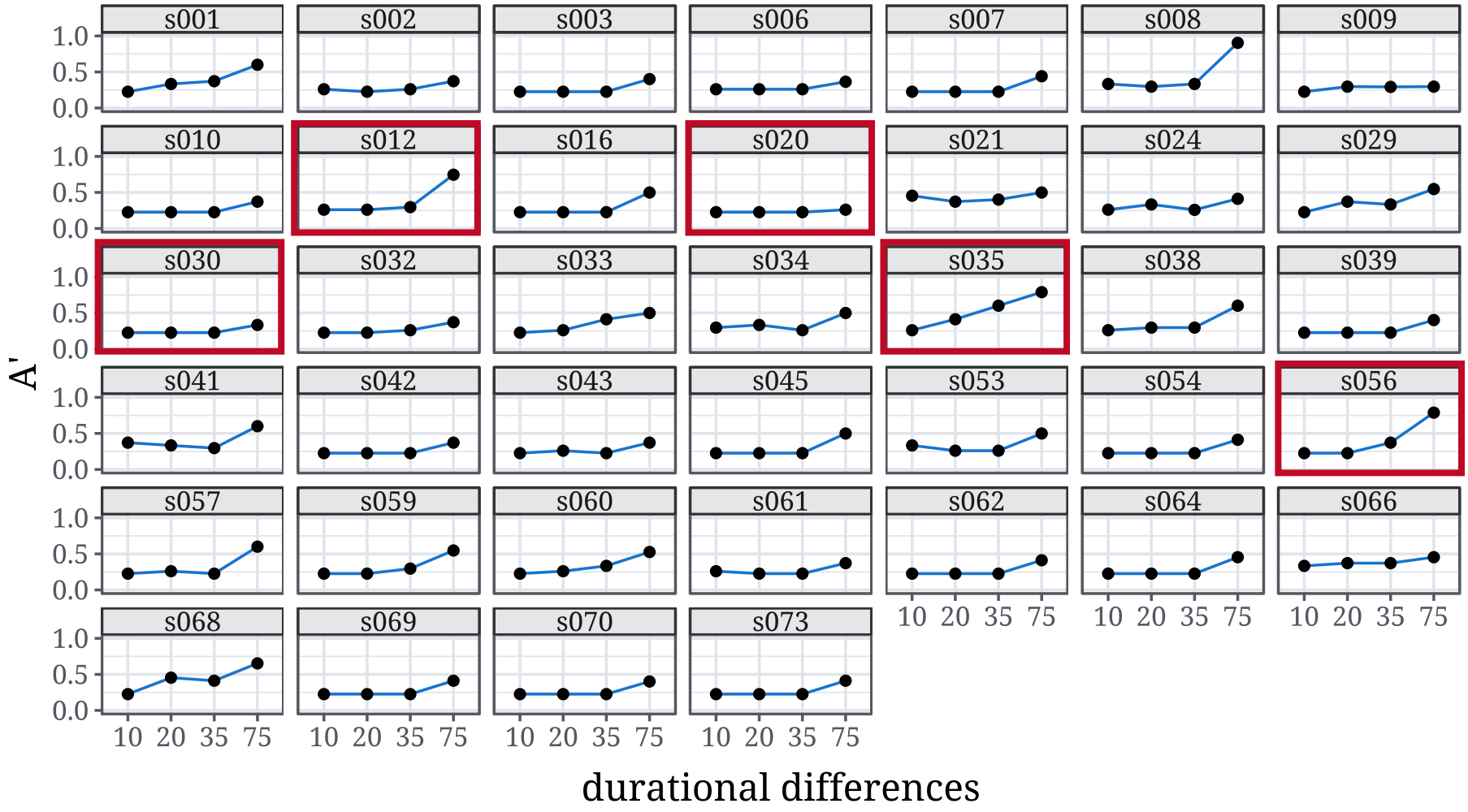
Understanding A′ sensitivity

- A′ is a measure from signal detection theory
- A′ = 0.5 means listeners are guessing
- A′ closer to 1 means listeners reliably detect the difference



Schmitz (2022)

Overall perceptual sensitivity across duration differences



Schmitz (2022)

Overall perceptual sensitivity across duration differences

- sensitivity (A') is lowest when the /s/ duration difference is very small
- listeners perform near chance for 10 ms and mostly also for 20 ms
- sensitivity rises sharply around 35 ms
- highest sensitivity for the 75 ms condition

Schmitz (2022)

Real words vs pseudowords

- both real words and pseudowords show the same threshold behaviour
- lexical familiarity does not eliminate the difficulty of perceiving small durations
- but real words sometimes show slightly higher sensitivity at larger differences
    - possible contribution of lexical expectations
- overall pattern remains essentially identical

Schmitz (2022)

Where does perception kick in?

- 35 ms appears to be the approximate threshold for reliable discrimination
- below 35 ms: performance stays at or near chance
- above 35 ms: sensitivity increases noticeably
- 75 ms is clearly perceptible for almost all listeners

Schmitz (2022)

Interpretation of findings

- what the results tell us about perception
  - listeners can perceive subphonemic /s/ duration differences
  - but only once differences exceed a certain perceptual threshold
  - extremely small cues (< 20–30 ms) are not perceptually accessible
  - aligns with known limits of temporal resolution in auditory perception
- production encodes more detail than perception can always exploit
- let's now connect these perceptual results to the broader theoretical landscape

Klatt & Cooper (1975); Schmitz (2022)

Why integration matters

- each domain gives only a partial view of linguistic structure
- **production** shows what speakers encode
- **perception** shows what listeners can access
- **representation models** explain what speakers and listeners learn
- integrating them reveals how the mental lexicon shapes fine phonetics

Production reflects learned form-meaning mappings

- English /s/ durations reflect morphological category
- German /s/ durations reflect similar category distinctions
- German a-schwa durations reflect semantic distinctions
- these patterns arise systematically, not accidentally
- suggests speakers *have* statistical patterns in the lexicon

Schmitz et al. (2021); Schmitz & Bear-Henney (2024); Schmitz (in prep.)

Lexical representation links perception and production

- discriminative models learn mappings between form and meaning
- fine phonetic detail emerges from lexical associations
- duration reflects how strongly a form predicts a meaning and vice versa
- the lexicon therefore shapes both
  - how speakers produce words
  - how listeners interpret the speech signal

Baayen et al. (2019)

Why emergentist models fit the data

- models do not discard subphonemic information early
- they store gradient associations across many experiences
- they naturally predict durational differences seen in production
- they allow perceptual thresholds to coexist with production detail
- they treat phonetics and morphology as interconnected learning outcomes

e.g. Baayen et al. (2019); Beach et al. (2021)

Why abstractionist models do not fit the data

- abstractionist models assume fast mapping to phonemes
- phonemic identity of /s/ is constant across items
- therefore, duration should not carry morphological information
- but production and perception patterns contradict this
- learners appear sensitive to gradient detail stored in experience
- in sum, early full abstraction would erase exactly the differences we observe

e.g. McClelland & Elman (1986); Norris & McQueen (2008)

What the combined evidence tells us

- the speech signal contains subtle morphological and semantic information
- speakers encode this information in subphonemic detail
- listeners detect some of this detail, depending on perceptual thresholds
- lexical representation links these two sides of the system

Schmitz (2022)

Implications for models of speech perception

- strict early abstraction cannot account for the perceptual threshold observed
- emergentist frameworks integrate gradient detail naturally
- perceptual thresholds fit well with usage-based learning
- perception operates on a mixture of acoustic and lexical cues
- subphonemic detail may inform lexical access in small but meaningful ways

Schmitz (2022)

What this teaches us about the mental lexicon

- the lexicon encodes probabilistic phonetic associations
- speakers and listeners adapt to fine distributional patterns
- linguistic categories emerge from experience, not predefined rules
- the mental lexicon is sensitive to both categorical and gradient structure

= the lexicon as a network shaped by experience

What we learned from the perception experiment

- listeners can perceive duration differences in word-final /s/
- perceptual sensitivity shows a clear threshold around ~35 ms
- very small differences (10–20 ms) are not reliably detectable
- real words and pseudowords show similar sensitivity patterns
- perception aligns with known limits of auditory temporal resolution
- perception is gradient, not categorical, but also not infinitely precise

Where this line of research is heading

- investigating perception of German a-schwa duration
- linking individual perceptual thresholds to lexical structure
- modelling how learners acquire subphonemic cues across development
- testing how subphonemic detail influences lexical access in real-time tasks
- integrating production and perception in larger discriminative models

Take-home message

- speakers encode subtle information in the speech signal
- listeners perceive these cues once they exceed perceptual thresholds
- the structure of the lexicon gives rise to gradient phonetic patterns
- subphonemic detail matters because it reflects how form and meaning interact in the lexicon and is, itself, a reflection of this interaction
- understanding subphonemic detail reshapes our models of speech processing

THANK YOU!

References 1

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 2019, 4895891.

Baer-Henney, D., & Schmitz, D. (submitted). *Homophones are not homophonous: The case of word-final /s/ in German*.

Beach, S. D., Ozernov-Palchik, O., May, S. C., Centanni, T. M., Gabrieli, J. D. E., & Pantazis, D. (2021). Neural decoding reveals concurrent phonemic and subphonemic representations of speech across tasks. *Neurobiology of Language*, *2*(2), 254–279.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper and Row.

Kiparsky, P. (1982). Lexical morphology and phonology. In I. Yang (Ed.), *Linguistics in the morning calm: Selected papers from SICOL1* (pp. 3–91). Hanshin.

Klatt, D. H., & Cooper, W. E. (1975). Perception of segment duration in sentence contexts. In A. Cohen & S. G. Nooteboom (Eds), *Structure and process in speech perception* (pp. 69–89). Springer.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(01), 1–75.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

References 2

Plag, I., Homann, J., & Kunter, G. (2017). Homophony and morphology: The acoustics of word-final S in English. *Journal of Linguistics*, *53*(1), 181–216.

Roelofs, A., & Ferreira, V. S. (2019). The architecture of speaking. In P. Hagoort (Ed.), *Human language: From genes and brains to behavior* (pp. 35–50). MIT Press.

Schmitz, D. (in prep.). *Autohyponymy and the phonetic signal: The duration of word-final -er in German*.

Schmitz, D. (2022). *Production, perception, and comprehension of subphonemic detail: Word-final /s/ in English*. (Studies in Laboratory Phonology 11). Language Science Press.

Schmitz, D., & Baer-Henney, D. (2024). Morphology renders homophonous segments phonetically different: Word-final /s/ in German. *Proceedings of Speech Prosody 2024*, 587–591.

Schmitz, D., Baer-Henney, D., & Plag, I. (2021). The duration of word-final /s/ differs across morphological categories in English: Evidence from pseudowords. *Phonetica*, *78*(5–6), 571–616.

Turk, A., & Shattuck-Hufnagel, S. (2020). *Speech timing*. Oxford University Press.

Zimmermann, J. (2016). Morphological status and acoustic realization: Findings from New Zealand English. In C. Carignan & M. D. Tyler (Eds), *Proceedings of the Sixteenth Australasian International Conference on Speech Science and Technology (SST-2016)* (pp. 201–204). ASSTA.