

The Dual Nature of Lengthening

Simon Betz, Petra Wagner
DiSS Workshop 2025

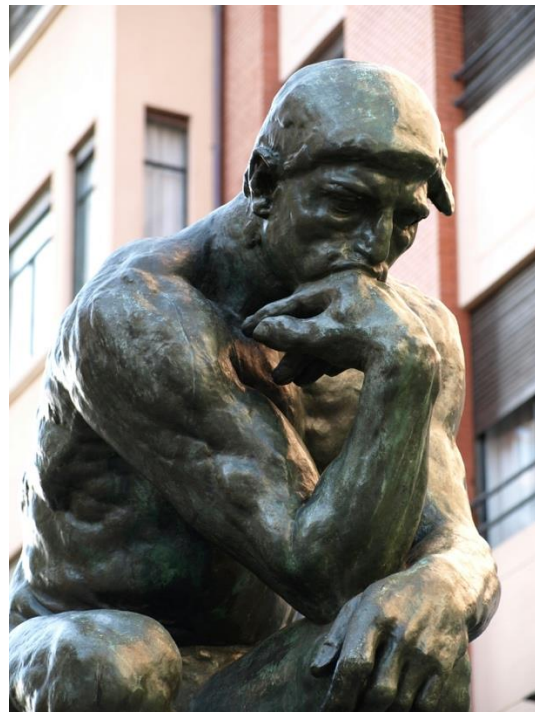
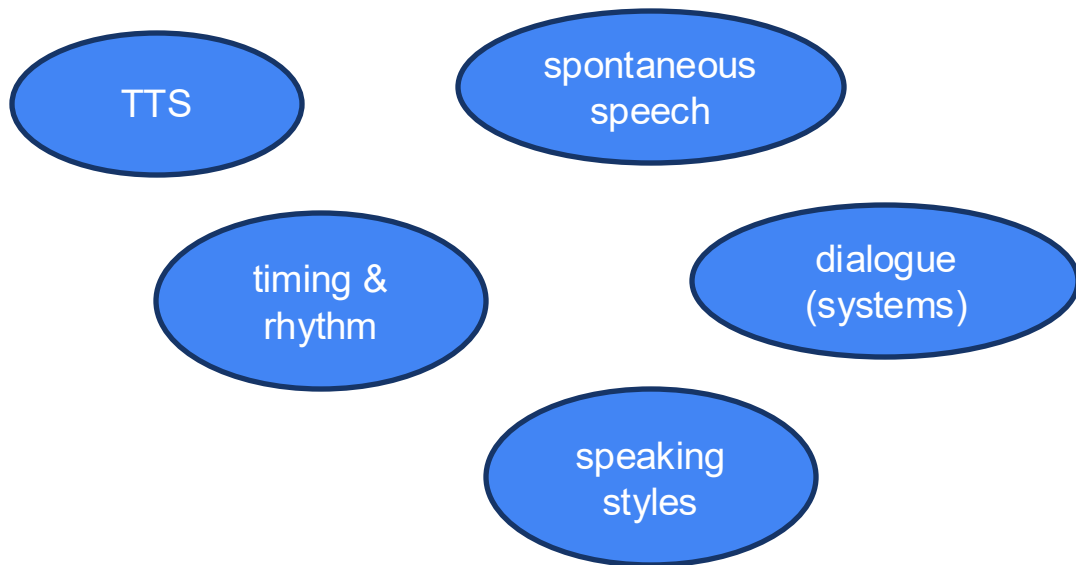
Public Service Announcement

All of our own papers mentioned in this presentation, and all slides, are available online and linked under this URL / QR code!

<https://length-eed2ab.pages.ub.uni-bielefeld.de/>



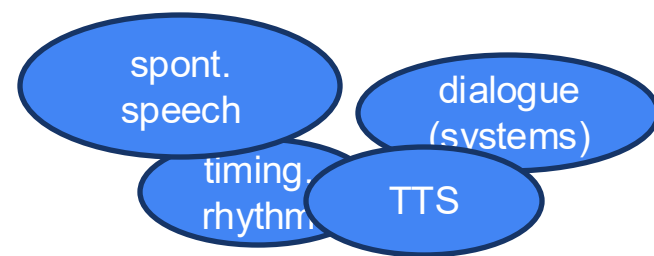
...a bunch of interests and ideas



Picture by Fernando Santando on Unsplash

... a bunch of interests and ideas

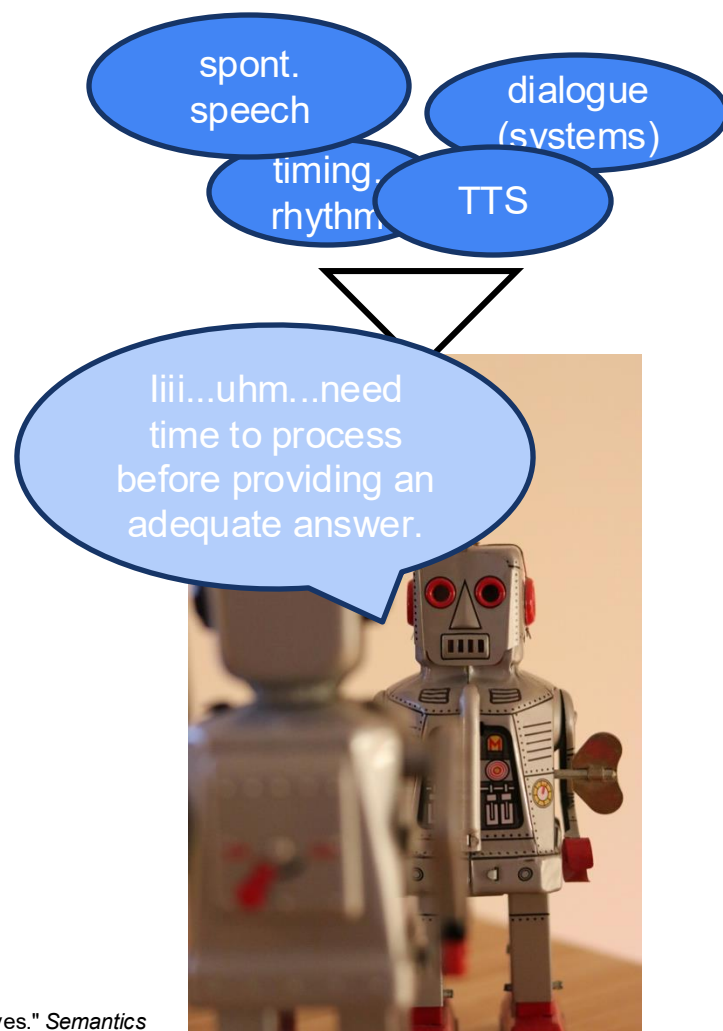
- Dialogue systems sometimes delay their answers...as do people
- These delays need to be produced or synthesized
- Hesitations as “forward-looking, time-buying disfluencies” might be part of the answer... (Ginzburg et al., 2014; Betz, 2020)



Picture by Theo on Unsplash

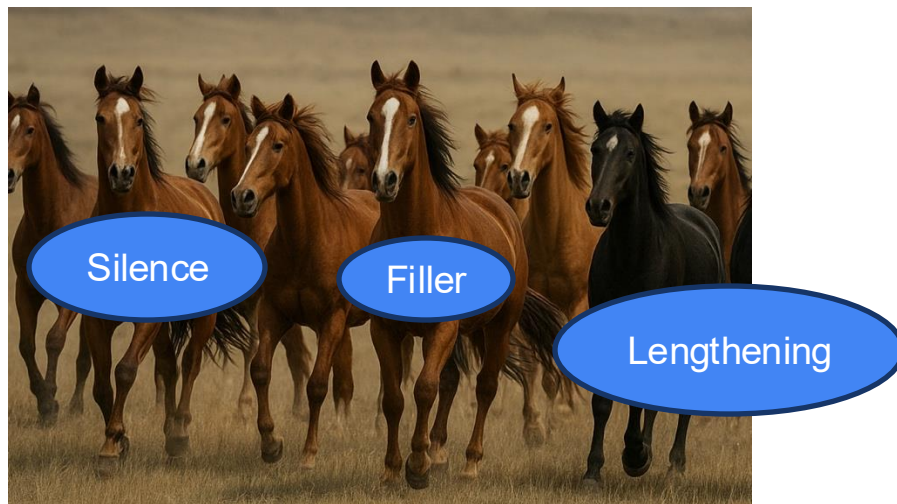
... a bunch of interests and ideas

- Dialogue systems sometimes delay their answers...as do people
- These delays need to be produced or synthesized
- Hesitations as “forward-looking, time-buying disfluencies” might be part of the answer... (Ginzburg et al., 2014; Betz, 2020)



The potential of hesitation lengthening

- Lengthening one of the most frequent disfluency phenomena, yet a “dark horse in the stable” (*Eklund 2001*)
- Often not the center of attention, unnoticed next to fillers and silences
- First hunch of a dual nature :
 - inserted strategically
 - often passes unnoticed



Hesitation vs. Non-Hesitation Lengthening

- Lengthening ubiquitous in languages: Phrase finality, domain initiality, accentuation, contrast, segmental quantity contrasts (gemination, vowel quantity)
- Non-hesitation lengthening tends to:
 - show other prosodic features (dynamic tonal contours, varied intensity etc.)
 - be not restricted to 1-2 phones
 - be syntagmatically restricted (stressed/stressable, final/initial)
 - be restricted in extent by phone elasticity (*Campbell et al., 1991*)

Hesitation vs. Non-Hesitation Lengthening

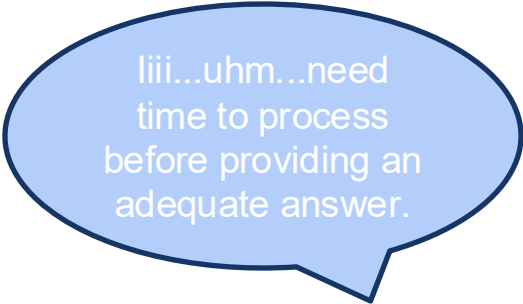
- Lengthening ubiquitous in languages: Phrase finality, domain initiality, accentuation, contrast, segmental quantity contrasts (gemination, vowel quantity)
- Hesitation lengthening tends to be:
 - ~~flat show other prosodic features (dynamic tonal contours, varied intensity)~~
 - ~~not~~ restricted to 1-2 phones (vowel + nasal)
 - occurring in unpredictable places ~~syntagmatically restricted (stressed/stressable, final/initial)~~
 - unusually long ~~restricted in extent by phone elasticity~~

Hesitation vs. Non-Hesitation Lengthening

- Lengthening ubiquitous in languages: Phrase finality, domain initiality, accentuation, contrast, segmental quantity contrasts (gemination, vowel quantity)
- Hesitation lengthening tends to be:
 - ~~flat show other prosodic features (dynamic tonal contours, varied intensity)~~
 - ~~not~~ restricted to 1-2 phones (vowel + nasal)
 - occurring in unpredictable places ~~syntagmatically restricted (stressed/stressable, final/initial)~~
 - unusually long ~~restricted in extent by phone elasticity~~
- Two types not straightforward to distinguish (e.g., phrase final hesitations)!

Hesitation vs. Non-Hesitation Lengthening

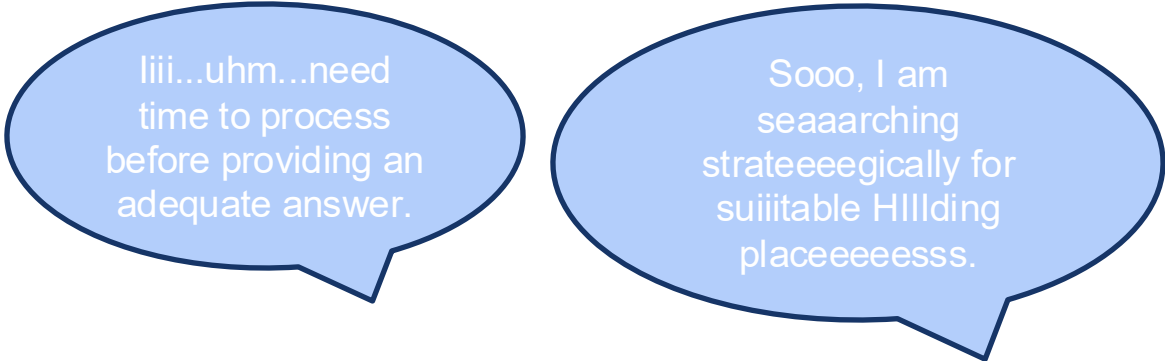
- Two types not straightforward to distinguish (e.g., phrase final hesitations)
- Second hunch of a dual nature:
 - inserted strategically
 - often passes unnoticed



liii...uhm...need
time to process
before providing an
adequate answer.

Hesitation vs. Non-Hesitation Lengthening

- Two types not straightforward to distinguish (e.g., phrase final hesitations)
- Second hunch of a dual nature:
 - inserted strategically (perhaps nearby accents or phrase boundaries?)
 - often passes unnoticed (because it is hidden?)



liii...uhm...need
time to process
before providing an
adequate answer.

Sooo, I am
seaaarching
strateeeegically for
suiiitable Hlllding
placeeeeeeesss.

Hesitation vs. Non-Hesitation Lengthening

- Two types not straightforward to distinguish (e.g., phrase final hesitations)
- Second hunch of a dual nature:
 - inserted strategically (perhaps nearby accents or phrase boundaries?)
 - often passes unnoticed (because it is hidden?)

How do we measure something hidden (by design)?



Lengthening: A Journey of Discovery

2015: Modular Disfluency Synthesis

2016: Disfluent Lengthening in Spontaneous Speech

2016: Synthesized lengthening of function words

2017: Semi-Automatic Classification

2018: Interactive Hesitation Synthesis

2019: The Greennnn Tree

2023: Cognitive Load and Multimodal Hesitation

2024: Multimodal Lengthening and Non-Understanding



Modular Disfluency Synthesis 2015

- Incremental processing enables real-time interactive speech synthesis.

MODULAR SYNTHESIS OF DISFLUENCIES FOR CONVERSATIONAL SPEECH SYSTEMS

Simon Betz, Petra Wagner, David Schlangen

Bielefeld University

Faculty of Linguistics and Literary Studies

Phonetics and Phonology Workgroup, Dialogue Systems Group

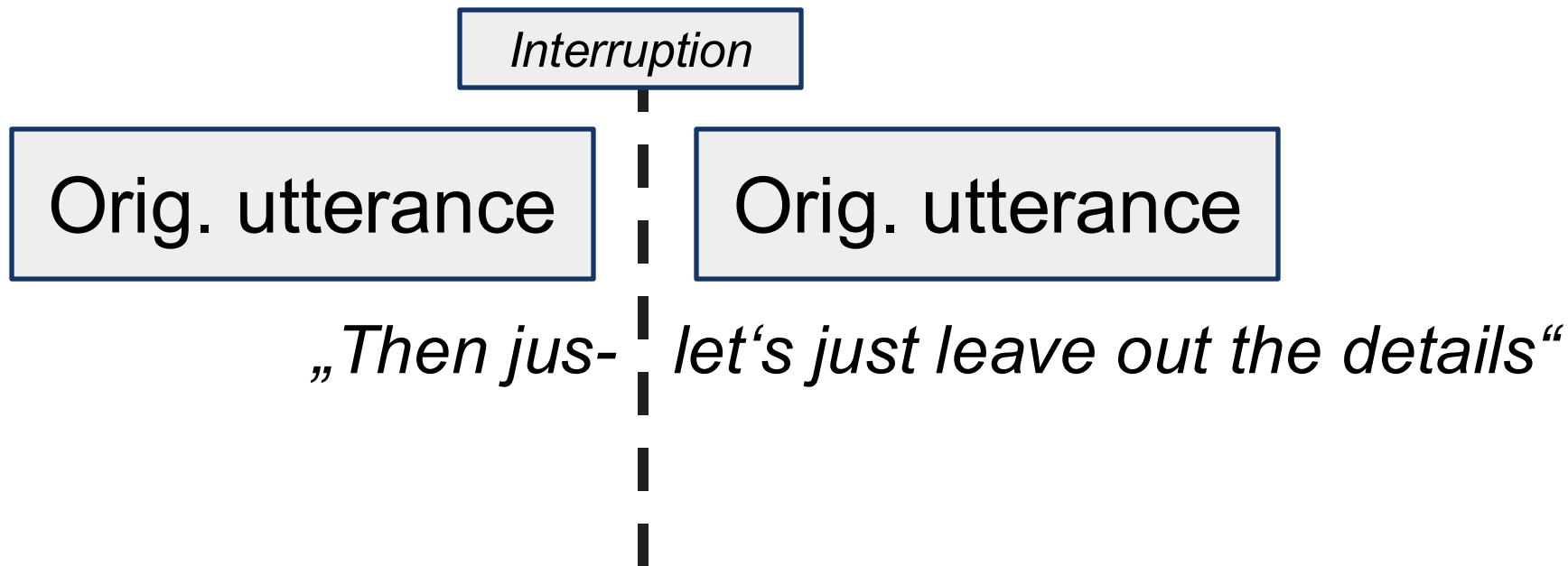
simon.betz@uni-bielefeld.de



Modular Disfluency Synthesis 2015

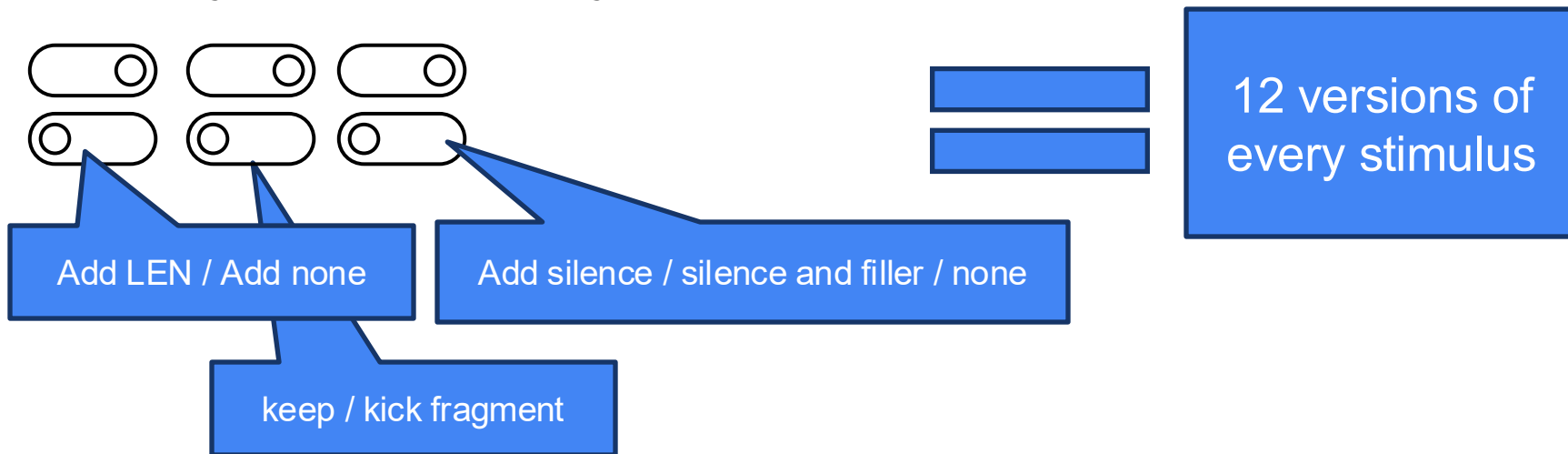
- Incremental processing risks committing to wrong input.
- No guarantee that it always works as quickly as desired.
- **This is precisely when speakers produce disfluencies.**
- As a first test, we synthesized different types of disfluencies and collected feedback on sound quality
- Goal: test which disfluencies work
- Hypothesis: disfluent dialogue systems will sound worse, but the gain in interactivity will make up for it

Modular Disfluency Synthesis 2015



Modular Disfluency Synthesis 2015

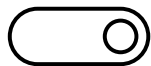
„Then jus- let’s just leave out the details“



Modular Disfluency Synthesis 2015

Example 100

„Then jus- let’s just leave out the details“



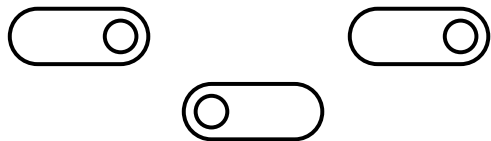
„Then::: let’s just leave out the details“

(Length. ON Fragg. OFF Silence OFF)

Modular Disfluency Synthesis 2015

Example 102

„Then jus- let’s just leave out the details“

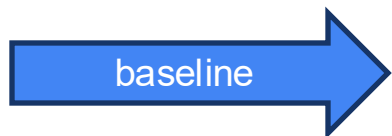


„Then::: ... uhm... let’s just leave out the details“

(Length. ON Fragn. OFF Silence ON+Filler)

Modular Disfluency Synthesis 2015: Results

- Fillers and fragments get low ratings.
- Lengthenings (and some silences) get very high ratings.
- Some configurations get even higher ratings than the fluent baseline!



Lengthening	Fragment	Silence	Filler	Rating
0	0	0	0	3.41
1	0	0	0	3.45
2	0	0	0	3.52

Take-Home-Message:

Lengthening is not the same as other disfluencies!

It seems to have more potential in terms of sound quality and acceptability.

Disfluent Lengthening in Spontaneous Speech 2016

- Goal: Use lengthening as core for incremental disfluency synthesis.
- More foundational research needed.
- Large-scale corpus study into durational properties of lengthening
 - ▲ Free spontaneous dialogue, German

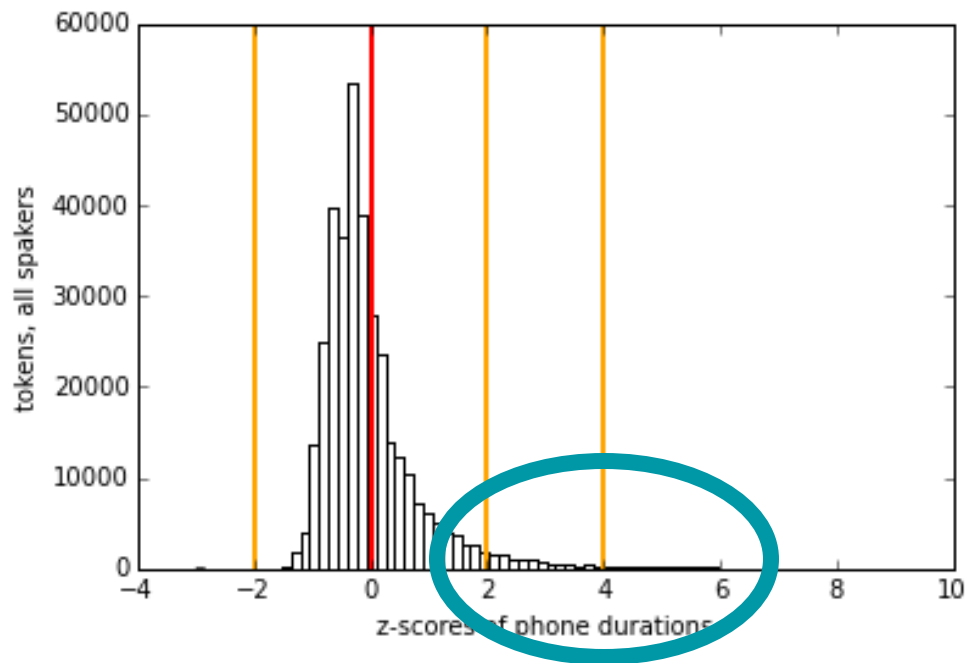
DISFLUENT LENGTHENING IN SPONTANEOUS SPEECH

Simon Betz, Petra Wagner

*Universität Bielefeld, Phonetics and Phonology Workgroup
simon.betz@uni-bielefeld.de*

Disfluent Lengthening in Spontaneous Speech 2016

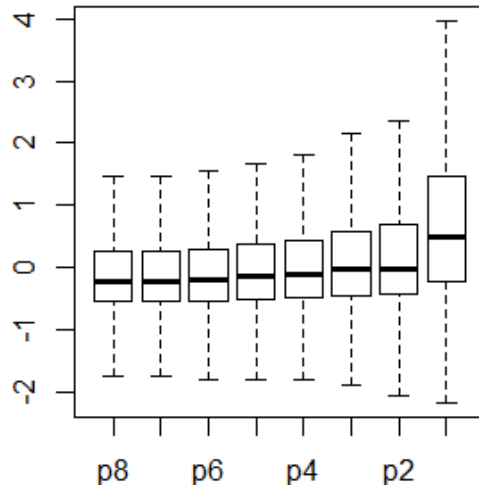
- How are phone durations distributed?
- Plot of all phone durations (no stops), z-normalized
- Why are there so many tokens above $z=2$?
- Almost all of them are **phrase final lengthening**.
- Hesitations hidden in the same duration range, but less frequent.



Disfluent Lengthening in Spontaneous Speech 2016

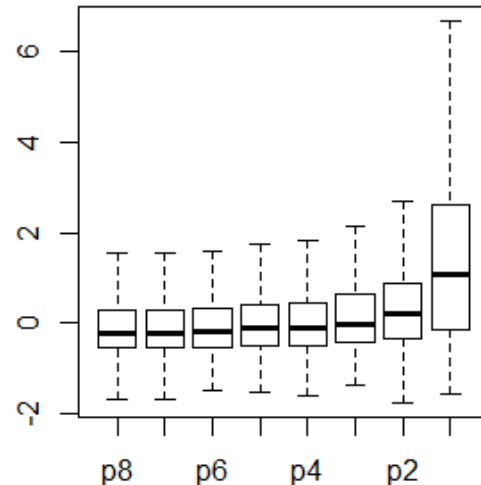
- gradual duration increase
- main load carried by last phone before boundary
- Same for lengthenings preceding fillers (uh, uhm)

The similarity between phrase-final and pre-filler lengthening suggests that the presence of a disfluency may split the intonation phrase into two.



Phrase final lengthening.

P8 = 8 phones before boundary
P1 = last phone before boundary
Y-axis: z-scores



Pre-filler lengthening.

Disfluent Lengthening in Spontaneous Speech 2016



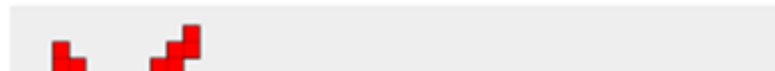
Where
lengthening

Take-Home-Message:

Hesitation lengthening can dress up as something else, e.g. as phrase-final lengthening. This makes it quite difficult to spot sometimes...

Synthesized lengthening of function words 2016

- Participants have to drag and



Synthesized lengthening of function words - The fuzzy boundary between fluency and disfluency

Simon Betz^{1,2,3}, Sina Zarriß^{2,3}, Petra Wagner^{1,3}

Bielefeld University, Bielefeld, Germany

¹Phonetics and Phonology Workgroup ²Dialogue Systems Group ³CITEC*

Fig. 1. Game scene with sound quality feedback buttons: very good, rather good, rather poor, very poor.

“Lengthening helps people to perform in tasks” 2016

- Two metrics:
- Sound quality
 - Generally good feedback, but:
 - Listeners don't like long lengthening
- Task performance
 - Longer lengthening = shorter relative task completion time

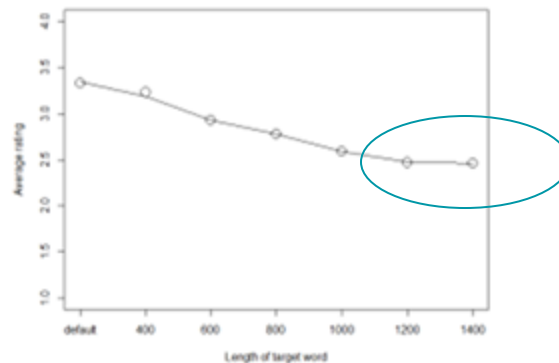


Fig. 2. User feedback with respect to word length. 4=good, 1=bad



Take-Home-Message:

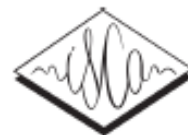
Listeners like hesitation lengthening, even synthesized, as long as it is not too long. Still, long lengthening positively affects performances in cognitive tasks. Probably due to the fact that it buys time for thinking.

Semi-Automatic Classification 2017

- *Lengthening is useful, we want more data.*
- A quite sobering glance at hesitation lengthening frequency in our two main corpora:

INTERSPEECH 2017

August 20–24, 2017, Stockholm, Sweden



Increasing Recall of Lengthening Detection via Semi-Automatic Classification

Simon Betz^{1,2}, Jana Voße¹, Sina Zarrieß^{1,2}, Petra Wagner^{1,2}

¹Bielefeld University, Bielefeld, Germany

²CITEC, Bielefeld, Germany

`simon.betz@uni-bielefeld.de`

Semi-Automatic Classification 2017

- Hypothesis: Lengthening is elusive
- Problem: How do we conduct corpus studies on a phenomenon that escapes the annotator?
- Solution: we built a simple lengthening detection tool
- Requires phonemically annotated corpus as input
- Outputs all instances with high z-normalized duration ($z \geq 3$)
- This provides flags for annotators
- All flags then manually* classified into hesitation / other

Take-Home-Message:

Lengthening buys extra dialogue time **without listeners noticing!**



Article

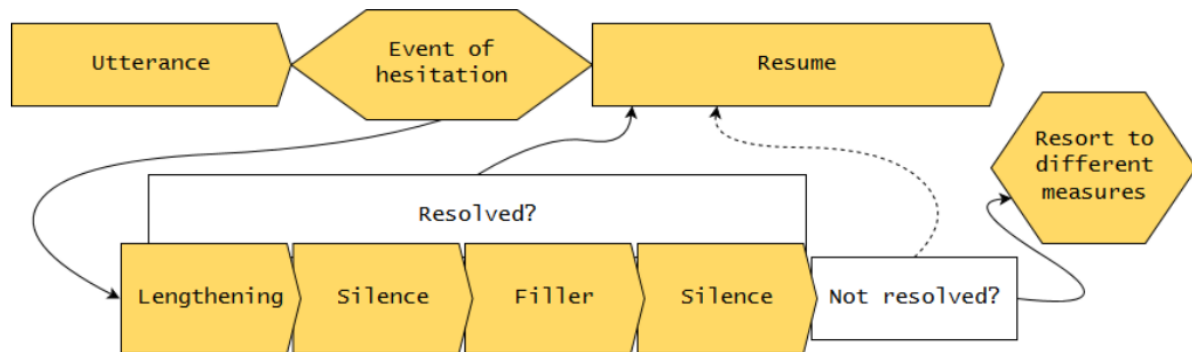
Interactive Hesitation Synthesis: Modelling and Evaluation

Simon Betz ^{1,2,3,*} , Birte Carlmeyer ^{1,3,4}, Petra Wagner ^{1,2} and Britta Wrede ^{1,4}

- ¹ Cluster of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, 33615 Bielefeld, Germany; b.carlmeyer@uni-bielefeld.de (B.C.); petra.wagner@uni-bielefeld.de (P.W.); bwrede@techfak.uni-bielefeld.de (B.W.)
 - ² Phonetics and Phonology Workgroup, Faculty of Linguistics and Literary Studies, Bielefeld University, 33615 Bielefeld, Germany
 - ³ Dialogue Systems Group, Faculty of Linguistics and Literary Studies, Bielefeld University, 33615 Bielefeld, Germany
 - ⁴ Applied Informatics Group, Faculty of Technology, Bielefeld University, 33615 Bielefeld, Germany
- * Correspondence: simon.betz@uni-bielefeld.de; Tel.: +49-521-106-3518

Interactive Hesitation Synthesis 2018

- *While hesitation.event: (defined and managed by the dialogue system)*
 - *Find the next best target* for lengthening*
 - *Apply as much lengthening as elasticity** allows*
 - *insert silence, cap at 1s*
 - *insert filler*
 - *insert silence, cap at 1s*
 - *give up ;)*

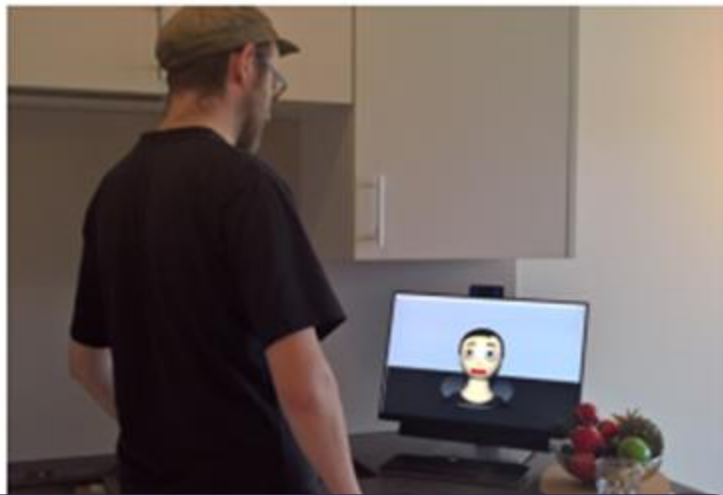


*Betz, Simon, Wagner, Petra, and Voße, Jana. "Deriving a strategy for synthesizing lengthening disfluencies based on spontaneous conversational speech data". *Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum*. Ed. Christoph Draxler and Felicitas Kleber. München: LMU, 2016. 19-22.

** Betz, Simon, Voße, Jana, and Wagner, Petra. "Phone Elasticity in Disfluent Contexts". *Fortschritte der Akustik - DAGA 2017*. Ed. Gerhard Schmidt, Bodo Nolte, Ulrich Heute, and Deutsche Gesellschaft für Akustik e.V. Berlin: Deutsche Gesellsch. f. Akustik, 2017. 1462-1464. 42

Interactive Hesitation Synthesis 2018

- Tested in an item-retrieval study
- Virtual agent on a screen gives instructions where to find items
- Event of Hesitation = lack of eye contact
- Two conditions: Hesitating and non-hesitating system
- **Participants instructed by**



Take-Home-Message:

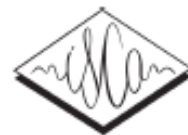
Lengthening works as entry point for hesitation insertion.
Real-time adaptive synthesis of hesitations possible.
Loss in sound quality trades off with gain in task performance.

The Greennnn Tree 2019

- For our hesitation insertion algorithm, we assumed that it is fine to lengthen at the next best position.
- But: it makes a difference, where in a word the lengthening is.

INTERSPEECH 2019

September 15–19, 2019, Graz, Austria



The greennn tree – lengthening position influences uncertainty perception

Simon Betz¹, Sina Zarrieß¹, Éva Székely², Petra Wagner¹

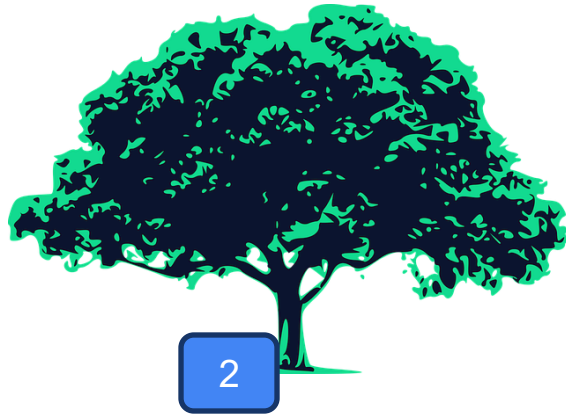
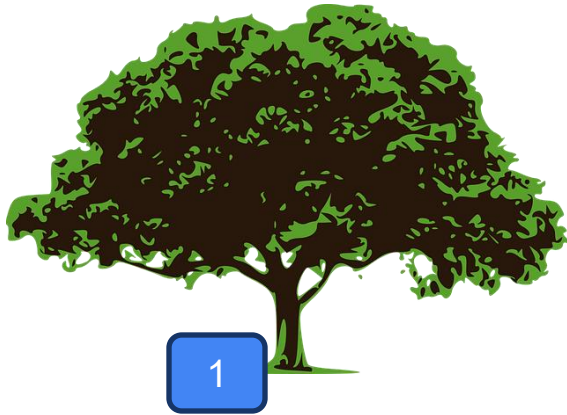
¹Faculty of Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany

²Division of Speech, Music and Hearing, KTH Royal Institute of Technology, Stockholm, Sweden

simon.betz@uni-bielefeld.de

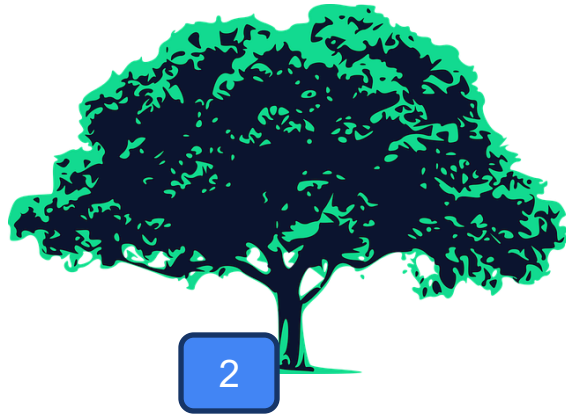
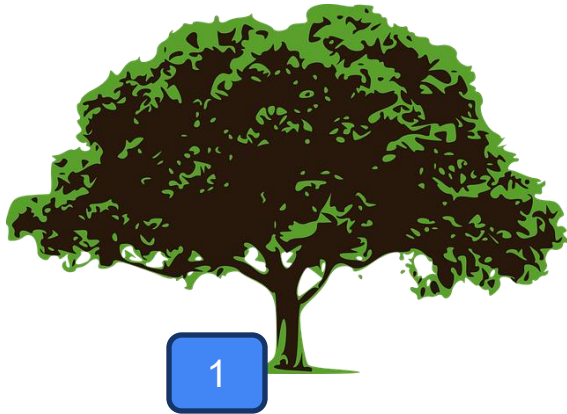
The Greennnnn Tree 2019

- Question for you: Which image am I referring to?
- „The green tree.“



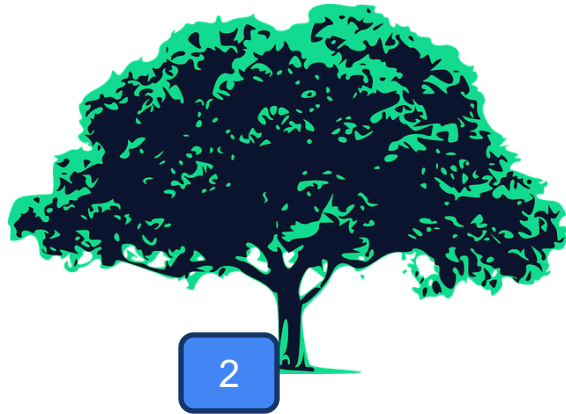
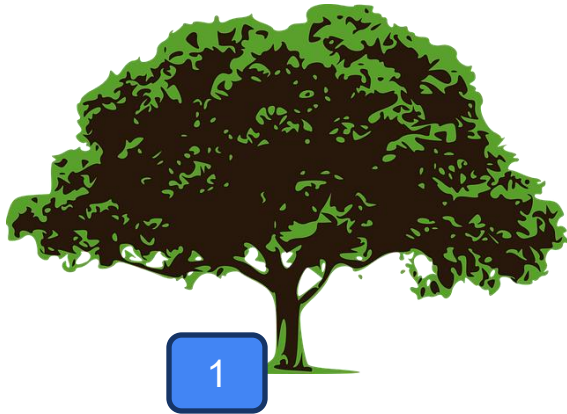
The Greennnnn Tree 2019

- Question for you: Which image am I referring to now?
- „The grrrrrr:een tree.“



The greennnn tree 2019

- Question for you: Which image am I referring to now?
- „The greennnn: tree.“



The Greennnnn Tree 2019

- Usual result: Initial lengthening = Uncertainty about word itself, Final lengthening = Uncertainty about following word



Green
Tree



Grrrreen
Tree.



Greennn
Tree.

The Greennnnn Tree 2019

- Perception study to test this intuition
- Pretext: Our system can signal uncertainty about colors and shapes.

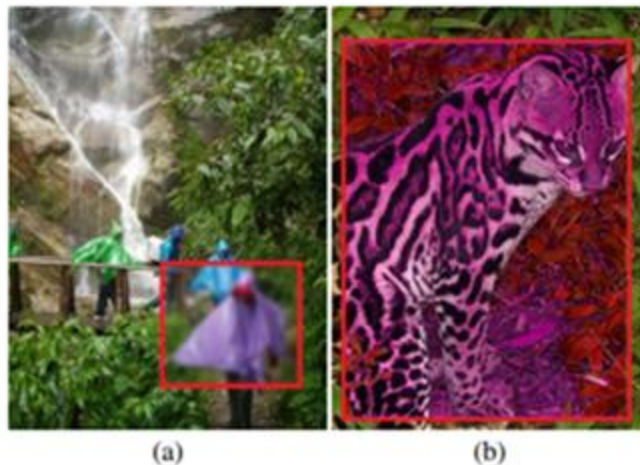


Figure 1: *Two images (shown as examples in the warm-up phases of our experiment) illustrating different domains of uncertainty: object category (a) and color (b).*

The Greennnn Tree 2019

- Stimuli modified to have either initial or final lengthening in a color word
- Participants have to rate the system's uncertainty wrt color and shape for each stimulus.

RESULTS:

Initial LEN = word itself

Final LEN = next word or word itself

No LEN = no uncertainty

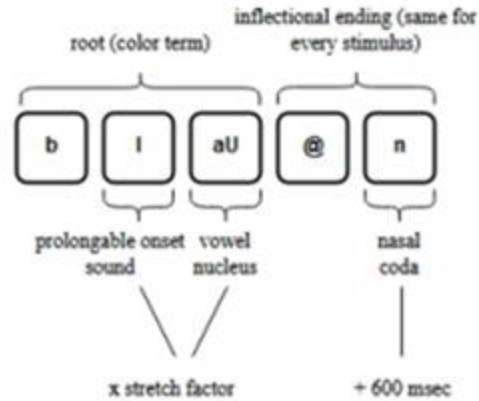


Figure 3: Duration modification for initial and final lengthening exemplified on the color term "blauen" (blue + ending).



Sie können die Audiowiedergabe durch Klick auf das Symbol oder durch Drücken der Leertaste starten.

Das System war in Bezug auf die Farbe unsicher.

sehr sicher sehr unsicher

Das System war in Bezug auf das Objekt unsicher.

sehr sicher sehr unsicher

Abbrechen

The Greennnnn Tree 2019

- *How many words would you need to phrase „grrrreen tree“ without lengthening available?*
- *„The tree that is actually green, not the one that is blueish green...“*
- For conveying uncertainty, lengthening is very efficient.
- One reason why lengthening is usually not rated negatively – it is familiar as part of fluent communication?

Take-Home-Message:

If lengthening is perceived, it can have massive communicative impact.

Cognitive Load and Multimodal Hesitation 2023

- We know that lengthening is very frequent, very useful in communication, and very hard to notice in corpora

Open Access

Article

Cognitive Load Increases Spoken and Gestural Hesitation Frequency

by Simon Betz *  , Nataliya Bryhadyr, Olcay Türk and Petra Wagner 

Department of Linguistics, Phonetics Workgroup, Bielefeld University, 33615 Bielefeld, Germany

* Author to whom correspondence should be addressed.

Languages **2023**, *8*(1), 71; <https://doi.org/10.3390/languages8010071>

Submission received: 19 November 2022 / Revised: 13 February 2023 / Accepted: 24 February 2023 /

Published: 2 March 2023

(This article belongs to the Special Issue **Pauses in Speech**)

Cognitive Load and Multimodal Hesitation 2023



German Vocab Help:

Spielen – play

Zahnräder – *Cogworks* (card name)

Generieren – generate

Energie – energy

Rot – red

Grün – green

Blau – blue

Nutzen – use

Schaltkreise – *Circuits* (card name)

Eins – one

Zwei – two

Drei – three

Filter – *Filter* (card name)

Karte – Card

Ziehen – draw

Jibjibjib – *jibjibjib* ;)

Cognitive Load and Multimodal Hesitation 2023

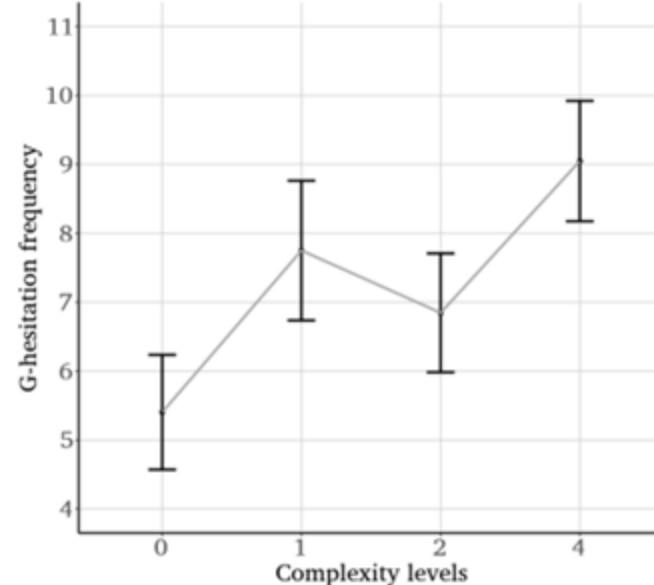
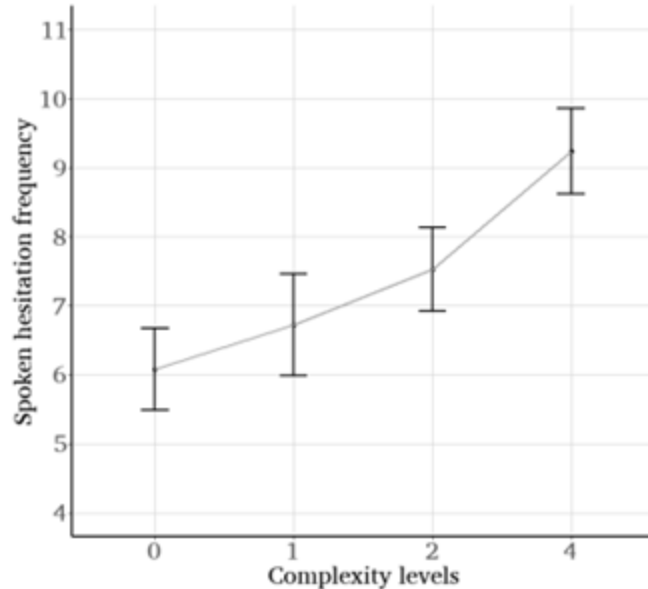
- Result 1: This setup yields abundant amounts of disfluencies

Hesitations	N	%	Corpus Specifics	
Silence	901	54.7	Words in corpus	8141
Lengthening	498	30.2	Corpus duration	110 min
Filler	249	15.1	Words affected by hesitation	20.2%
total	1648		Hesitations per minute	14.9

Corpus	Type	Size	Len. Freq.
DUEL	task-oriented	4.5 h	1.6 / minute
GECO	spontaneous	45 h	0.5 / minute
CARDGAME	task-oriented	1.9 h	4.5 / minute

Cognitive Load and Multimodal Hesitation 2023

- Result 2: Level of **complexity** (=cognitive load) is linearly correlated with **spoken** and **gestural** hesitation frequency



Cognitive Load and Multimodal Hesitation 2023

- Setup highly **efficient** to elicit ecologically valid disfluencies
- Spoken and gestural hesitations can be easily studied in connection
- Lengthening is very frequent.
- But more frequent than fillers? Interesting, and unexpected...
- *DiSS topic 2025: The nature of the corpus and the research question may influence how many lengthenings one finds.*

Cognitive Load and Multimodal Hesitation 2023



Take-Home-Message:

Lengthening is actually very frequent.
There is also lengthening in gestures.

Multimodal Lengthening and Non-Understanding



≡ Article Navigation

Predictability of Understanding in Explanatory Interactions Based on Multimodal Cues

[Olcay Türk](#), Faculty of Linguistics and Literary Studies, Bielefeld University, Germany, olcay.tuerk@uni-bielefeld.de

[Stefan Lazarov](#), Faculty of Arts and Humanities, Paderborn University, Germany, stefan.lazarov@upb.de

[Yu Wang](#), Faculty of Linguistics and Literary Studies, Bielefeld University, Germany, y.wang@uni-bielefeld.de

[Hendrik Buschmeier](#), Faculty of Linguistics and Literary Studies, Bielefeld University, Germany, hbuschme@uni-bielefeld.de

[Angela Grimminger](#), Faculty of Arts and Humanities, Paderborn University, Germany, angela.grimminger@uni-paderborn.de

[Petra Wagner](#), Faculty of Linguistics and Literary Studies, Bielefeld University, Germany, petra.wagner@uni-bielefeld.de

DOI: <https://doi.org/10.1145/3678957.3685741>

ICMI '24: [INTERNATIONAL CONFERENCE ON MULTIMODAL INTERACTION](#), San Jose, Costa Rica, November 2024

Multimodal Lengthening and Non-Understanding

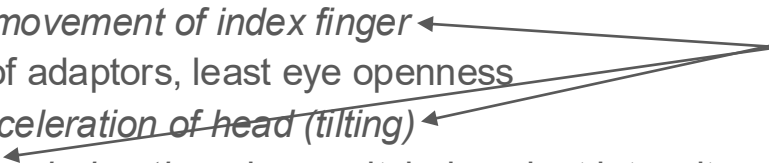
- Data: 4s windows of (non-)understanding in 21 explanatory dialogues
- RF classifier, ranking of predictors



Multimodal Lengthening and Non-Understanding

- Predictors: acoustics, multimodal kinematics and spatial info
- Non-understanding signalled by:
 - continued movement of index finger
 - Presence of adaptors, least eye openness
 - Lacking acceleration of head (tilting)
 - Longer durations, lower pitch, invariant intensity

Multimodal Lengthening and Non-Understanding

- Predictors: acoustics, multimodal kinematics and spatial info
 - Non-understanding signalled by:
 - *continued movement of index finger*
 - Presence of adaptors, least eye openness
 - *Lacking acceleration of head (tilting)*
 - *Longer speech durations*, lower pitch, invariant intensity
- cross-modal lengthening!
- 

Dual-Nature-Take-Home-Message:

Lengthenings are heavily multimodal and occur in places where speakers report a lack in understanding. These are *sometimes* perceived by an interlocutor.

Summary

- Lengthening is not a dark horse anymore - its talents have been revealed!
- Lengthenings are frequent and multimodal
- Lengthenings are harder to trace perceptually than other disfluencies
- Due to their **dual nature**, they can be used **to hide or signal** something: uncertainty, cognitive load, lack of understanding,...
- Lengthenings can also be used for:
 - buying dialogue time (like all hesitations)
 - buying dialogue time without the listener noticing
 - serve as the entry point for all hesitation intervals

Thank you!

Scan me for a
compilation of paper
links!



Future and Ongoing Work

- Replicate the green tree (with broader context)
- Rocket science (Nickname for a study to disentangle filler and lengthening signaling functions)
- Hesitation grounding (after a filler has been produced, longer silences seem tolerable)
- EMA (Articulography research on Hesitations)
- Role of Lengthening in Multimodal Ensembles

EMA

- ElectroMagnetic Articulography investigation of hesitation (Joint work with Malte Belz & Tine Mooshammer, in preparation)
- Investigating articulatory resting positions and their interaction with hesitations and pauses

