Carl von Ossietzky
Universität
Oldenburg

UNIVERSITÄT
BIELEFELD

Vocalic signals are worth strengthening, but consonants may not be

Motoki Saito & Jessie S. Nixon

25.06.2025
Workshop “Instability of acoustic-phonetic cues”
The 6th Phonetics and Phonology in Europe (PaPE 2025)
Palma, Spain

Phonetic realizations

- Segments can be different from token to token.
  e.g., duration, pitch, formants, intensity, etc.
  → Phonetic realizations/variations.

Morphological effects on phonetic realizations

- Segments can be realized differently accordingly to their morphological status.
  e.g., *lap**s*** [læps] vs. *lap**se*** [læps]

[1] Plag et al. (2017) and Schmitz and Baer-Henney (2024)
[2] Hay (2007) and Saito et al. (2024)
[3] Engemann (2023) and Song et al. (2013)

Morphological effects on phonetic realizations

- Segments can be realized differently accordingly to their morphological status.
  e.g., *lap**s*** [læps] vs. *lap**se*** [læps]
- Affixes (e.g., *lap**s***) are longer than pseudo-affixes (e.g., *lap**se***) [1]
  → Possibly because affixes have their own meanings [2]

[1] Plag et al. (2017) and Schmitz and Baer-Henney (2024)
[2] Hay (2007) and Saito et al. (2024)
[3] Engemann (2023) and Song et al. (2013)

Morphological effects on phonetic realizations

- Segments can be realized differently accordingly to their morphological status.
  e.g., *lap**s*** [læps] vs. *lap**se*** [læps]
- Affixes (e.g., *lap**s***) are longer than pseudo-affixes (e.g., *lap**se***) [1]
  → Possibly because affixes have their own meanings [2]
- No such effects have also been reported [3]

[1] Plag et al. (2017) and Schmitz and Baer-Henney (2024)
[2] Hay (2007) and Saito et al. (2024)
[3] Engemann (2023) and Song et al. (2013)

Phonetic nature of morphological effects?

- Morphological effects on phonetics...

↓

- They may be modulated by phonological/phonetic characteristics.

Association of morphological effects and vocality[4]

(Contains a vowel)	Reduction	No diff.	Enhancement
Yes	0	5	13
No	6	6	7

[4] Ben Hedia and Plag (2017), Engemann (2023), Hay (2007), Li et al. (2020), Plag and Ben Hedia (2018), Plag et al. (2017), Schmitz and Baer-Henney (2024), Schmitz et al. (2021), Seyfarth et al. (2017), Smith et al. (2012), Song et al. (2013), Sproat and Fujimura (1993), Strycharczuk and Scobbie (2016), Sugahara and Turk (2009), Walsh and Parker (1983), Zimmermann (2016), and Zuraw et al. (2021)

Association of morphological effects and vocality[4]

(Contains a vowel)	Reduction	No diff.	Enhancement
Yes	0	5	13
No	6	6	7

- Presence of a vowel and kinds of findings are associated significantly ($\chi^2(2, n = 37) = 7.870, p < .05$).

[4] Ben Hedia and Plag (2017), Engemann (2023), Hay (2007), Li et al. (2020), Plag and Ben Hedia (2018), Plag et al. (2017), Schmitz and Baer-Henney (2024), Schmitz et al. (2021), Seyfarth et al. (2017), Smith et al. (2012), Song et al. (2013), Sproat and Fujimura (1993), Strycharczuk and Scobbie (2016), Sugahara and Turk (2009), Walsh and Parker (1983), Zimmermann (2016), and Zuraw et al. (2021)

Vowels' advantages over consonants

Greater informativity

- vowels masked in English sentences $\rightarrow$ 34-49% of word identification accuracy.
- consonants masked in English sentences $\rightarrow$ 74-87% of word identification accuracy.

Better audibility

- With increased loudness, the difference in word identification accuracy decreases.

Vowels' advantages over consonants

Greater informativity

- vowels masked in English sentences $\rightarrow$ 34-49% of word identification accuracy.
- consonants masked in English sentences $\rightarrow$ 74-87% of word identification accuracy.

Better audibility

- With increased loudness, the difference in word identification accuracy decreases.

$\Downarrow$

If the speaker must spend more articulatory effort, they may do so more for vowels.

- so that their additional effort can be more effective and efficient.

Hypothesis

- Enhancement effects are tied to vowels, while they are not to consonants.

Data: Target suffixes

- Word-final *-t* [-t]
  - e.g., **Suffixal**: *mach**t*** [maxt] “makes”
  - e.g., **Non-suffixal**: *Wel**t*** [vɛlt] “world”

- Word-final *-er* [-ɐ]
  - e.g., **Suffixal**: *laut**er*** [laʊ̯tə] “louder”
  - e.g., **Non-suffixal**: *Vat**er*** [fɑːtə] “father”
    - * Excluded: Articles and pronouns.

Structure of the talk

1. Acoustic analysis → Study 1
   - Data
   - Analysis
   - Results
2. Articulatory analysis → Study 2
   - Data
   - Analysis
   - Results
3. Summary

Study 1: Data

- All the words with either *-t* or *-er* were collected from Karl Eberhards Corpus (KEC)[5].
- KEC:
  - German spontaneous speech.
  - 39 speakers.
  - Approximately 35 hours of speech.

	Morphemic	Non-morphemic
-er	2361 (471)	14502 (262)
-t	16030 (1406)	25869 (675)

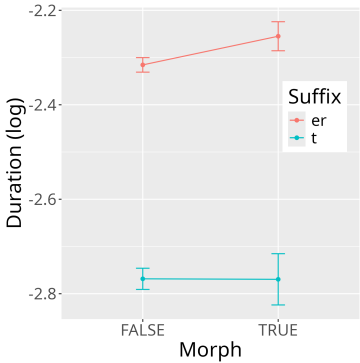
[5] Arnold and Tomaschek (2016)

Study 1: Analysis

Effects on duration

- `Suffix`: *-er* vs. *-t*.
- `Morph`: Non-morphemic vs. morphemic.
- `Speaker`: Speaker differences (as random intercept)
- and control variables such as frequency, speech rate, etc.

Study 1: Results



- Morphemic *-er* is significantly longer than non-morphemic *-er*.
- Morphemic *-t* is not significantly different from non-morphemic *-t*.

Interim summary

- Morphemic *-er* is longer in duration than non-morphemic *-er*.
- Morphemic and non-morphemic *-t* are the same in duration.

Interim summary

- Morphemic *-er* is longer in duration than non-morphemic *-er*.
- Morphemic and non-morphemic *-t* are the same in duration.

⇓

We found evidence of morphological effects on duration **only for vowels**.

To Study 2

- Study 1: Acoustic analysis
- **Study 2: Articulatory analysis**

Study 2: Data for the articulatory analysis

- Articulography part of KEC.
  - 13 speakers
  - Tongue tip movements
  - Electromagnetic articulography (EMA).

  → approximately 2 hours of speech.

  - Data points:

	Morphemic	Non-morphemic
-er	6280	40234
-t	30474	46782

Study 2: Analysis

Effects on vertical tongue positions

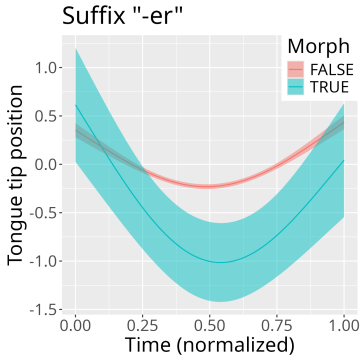
- `Time`: Time during the segment for the tongue position.
- `Suffix`: *-er* vs. *-t*.
- `Morph`: Non-morphemic vs. morphemic.
- and other control predictors and random effects.

Study 2: Analysis

Effects on vertical tongue positions

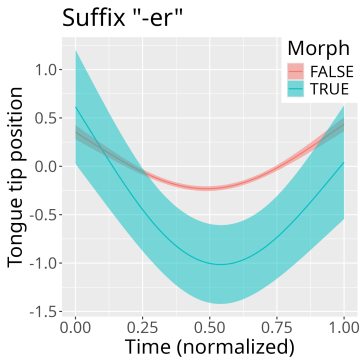
- `Time`: Time during the segment for the tongue position.
- ~~`Suffix`: *-er* vs. *-t*.~~ → Separate models
- `Morph`: Non-morphemic vs. morphemic.
- and other control predictors and random effects.

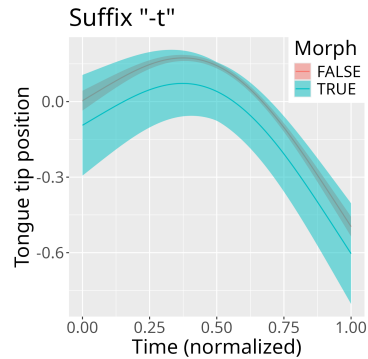
Study 2: Results



- Morphemic *-er* is significantly lower (more clearly articulated) than non-morphemic.
  - The differences are mainly manifested in the middle of the segment.

Study 2: Results





- Morphemic *-er* is significantly lower (more clearly articulated) than non-morphemic.
  - The differences are mainly manifested in the middle of the segment.
- No difference between morphemic and non-morphemic *-t*.

Summary of the observations

Morphemic *-er*:

- Longer in duration (Study 1)
- Clearer articulation (Study 2)

Morphemic *-t*:

- No difference in duration (Study 1)
- No difference in articulation (Study 2)

Discussion

Enhancement by morphology is modulated by the vowel-consonant contrast.

→ Why?

→ What is the difference between vowels and consonants?

Discussion

Enhancement by morphology is modulated by the vowel-consonant contrast.

→ Why?

→ What is the difference between vowels and consonants?

Our tentative proposal:

→ Speaker enhances the signal that is the **most informative** for the listener.

Possible answers

1. For communication effectiveness:
   - the speaker needs to maximize informativity/audibility.
   - the speaker tries to minimize articulatory effort [6].

[6] Lindblom (1983) and Nelson (1983)
[7] Clements (2009), Cole et al. (1996), and Kewley-Port et al. (2007)

Possible answers

1. For communication effectiveness:
   - the speaker needs to maximize informativity/audibility.
   - the speaker tries to minimize articulatory effort [6].

2. Vowels seem to be more informative/audibility than consonants [7].
   - Vowels masked $\rightarrow$ 34-49% of word identification accuracy
   - Consonants masked $\rightarrow$ 74-87% of word identification accuracy

[6] Lindblom (1983) and Nelson (1983)
[7] Clements (2009), Cole et al. (1996), and Kewley-Port et al. (2007)

Possible answers

1. For communication effectiveness:
   - the speaker needs to maximize informativity/audibility.
   - the speaker tries to minimize articulatory effort [6].

2. Vowels seem to be more informative/audibility than consonants [7].
   - Vowels masked $\rightarrow$ 34-49% of word identification accuracy
   - Consonants masked $\rightarrow$ 74-87% of word identification accuracy

$\Downarrow$

It may be wiser to spend effort on those that can readily be made salient.

$\rightarrow$ Namely those with better audibility.

[6] Lindblom (1983) and Nelson (1983)
[7] Clements (2009), Cole et al. (1996), and Kewley-Port et al. (2007)

For *-er* and *-t*

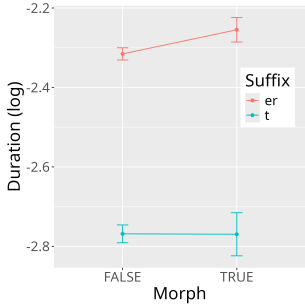
e.g., *-er* [ɐ] → Easily more audible by lengthening.

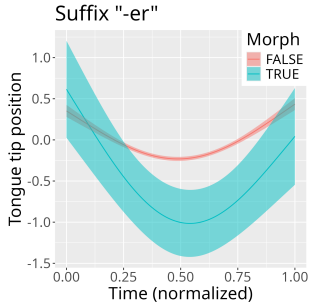
e.g., *-t* [-t] → Difficult to improve audibility.

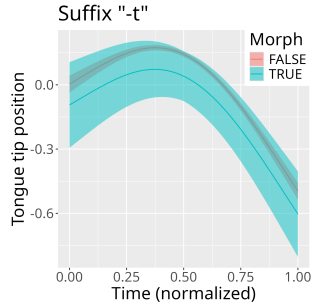
Summary

- Morphological effects may be modulated by phonetic properties.
- The speaker may enhance those signals that are the most beneficial for the listener.

Thank you very much!







Acknowledgments

This study and this talk were supported and enabled by

- R. Harald Baayen (University of Tübingen)
- Esther Ruigendijk (University of Oldenburg)
- The faculty III Dutch study department of University of Oldenburg.

References I

[1] Arnold, D., & Tomaschek, F. (2016).The Karl Eberhards Corpus of spontaneously spoken southern German in dialogues — audio and articulatory recordings. Tagungsband der 12. Tagung Phonetik und Phonologie im deutschsprachigen Raum, 9–11.

[2] Ben Hedia, S., & Plag, I. (2017).Gemination and degemination in English prefixation: Phonetic evidence for morphological organization. Journal of Phonetics, 62, 34–49. https://doi.org/10.1016/j.wocn.2017.02.002

[3] Clements, G. N. (2009). Does sonority have a phonetic basis? In Contemporary views on architecture and representations in phonology. The MIT Press. https://doi.org/10.7551/mitpress/9780262182706.003.0007

[4] Cole, R. A., Yan, Y., Mak, B., Fanty, M., & Bailey, T. (1996).The contribution of consonants versus vowels to word recognition in fluent speech. 1996 IEEE International Conference on Acoustics, Speech, and Signal Processing Confe 2, 853–856. https://doi.org/10.1109/ICASSP.1996.543255

[5] Engemann, M. (2023). Paradigm uniformity in inflectional stems: Durational differences in production and percep Düsseldorf university press. https://doi.org/doi:10.1515/9783111017754

References II

[6] Faaß, G., & Eckart, K. (2013). SdeWaC: A corpus of parsable sentences from the web. In I. Gurevych, C. Biemann, & T. Zesch (Eds.), Language processing and knowledge in the web (pp. 61–68). Springer.

[7] Hay, J. (2007). The phonetics of 'un'. In J. Munat (Ed.), Lexical creativity, texts and contexts (pp. 39–57). John Benjamins.

[8] Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007).Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. The Journal of the Acoustical Society of America, 122(4), 2365–2375. https://doi.org/10.1121/1.2773986

[9] Li, V. G., Oh, S., Chopra, G., Celli, J., & Shaw, J. A. (2020).Articulatory correlates of morpheme boundaries: Preliminary evidence from intra- and inter-gestural timing in the articulation of the English past tense. Proceedings of ISSP 2020 - 12th International Seminar on Speech Production.

[10] Lindblom, B. (1983). Economy of speech gestures. In P. F. MacNeilage (Ed.), The production of speech (pp. 217–245). Springer-Verlag.

[11] MacKenzie, L., & Turton, D. (2020).Assessing the accuracy of existing forced alignment software on varieties of british english. Linguistics Vanguard, 6(s1), 20180061.

References III

[12] Nelson, W. L. (1983).Physical principles for economies of skilled movements. Biological Cybernetics, 46(2), 135–147. https://doi.org/10.1007/BF00339982

[13] Plag, I., & Ben Hedia, S. (2018). The phonetics of newly derived words: Testing the effect of morphological segmentability on affix duration. In S. Arndt-Lappe, A. Braun, C. Moulin, & E. Winter-Froemel (Eds.), Expanding the lexicon: Linguistic innovation, morphological productivity, and ludicity (pp. 93–116). De Gruyter. https://doi.org/10.1515/9783110501933-095

[14] 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. https://doi.org/10.1017/S0022226715000183

[15] Saito, M., Tomaschek, F., Sun, C.-C., & Baayen, R. H. (2024). Articulatory effects of frequency modulated by semantics. In M. Schlechtweg (Ed.), Interfaces of phonetics (pp. 125–154). De Gruyter Mouton. https://doi.org/doi:10.1515/9783110783452-005

[16] Schmitz, D., & Baer-Henney, D. (2024).Morphology renders homophonous segments phonetically different: Word-final /s/ in German. Proceedings of Speech Prosody 2024 (SP2024), 587–591.

References IV

[17] 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. https://doi.org/10.1515/phon-2021-2013

[18] Seyfarth, S., Garellek, M., Gillingham, G., Ackerman, F., & Malouf, R. (2017).Acoustic differences in morphologically-distinct homophones. Language, Cognition and Neuroscience, 33(1), 32–49. https://doi.org/10.1080/23273798.2017.1359634

[19] Smith, R., Baker, R., & Hawkins, S. (2012).Phonetic detail that distinguishes prefixed from pseudo-prefixed words. Journal of Phonetics, 40(5), 689–705. https://doi.org/10.1016/j.wocn.2012.04.002

[20] Song, J. Y., Demuth, K., Shattuck-Hufnagel, S., & Ménard, L. (2013).The effects of coarticulation and morphological complexity on the production of English coda clusters: Acoustic and articulatory evidence from 2-year-olds and adults using ultrasound. Journal of Phonetics, 41(3-4), 281–295.

[21] Sproat, R., & Fujimura, O. (1993).Allophonic variation in English /l/ and its implications for phonetic implementation. Journal of Phonetics, 21(3), 291–311. https://doi.org/10.1016/s0095-4470(19)31340-3

References V

[22] Strycharczuk, P., & Scobbie, J. M. (2016).Gradual or abrupt? The phonetic path to morphologisation. Journal of Phonetics, 59, 76–91. https://doi.org/10.1016/j.wocn.2016.09.003

[23] Sugahara, M., & Turk, A. (2009).Durational correlates of English sublexical constituent structure. Phonology, 26(3), 477–524.

[24] Walsh, T., & Parker, F. (1983).The duration of morphemic and non-morphemic /s/ in English. Journal of Phonetics, 11(2), 201–206.

[25] Zimmermann, J. (2016).Morphological status and acoustic realization: Findings from New Zealand English. Proceedings of the Sixteenth Australasian International Conference on Speech Science a (December), 201–204.

[26] Zuraw, K., Lin, I., Yang, M., & Peperkamp, S. (2021).Competition between whole-word and decomposed representations of English prefixed words. Morphology, 31(2), 201–237. https://doi.org/10.1007/s11525-020-09354-6

Covariates and factors in Study 1

- `Freq`: Word frequency (log) from the SdeWaC corpus [8].
- `UttInitial`: Whether the word is utterance-initial.
- `UttFinal`: Whether the word is utterance-final.
- `uDur`: Duration of the utterance.
- `uNumSyl`: The number of syllables in the utterance.
- `uSpRate`: Utterance-based speech rate.
- `wDur`: Duration of the word.
- `wNumSyl`: The number of syllables in the word.
- `wSpRate`: Word-based speech rate.

[8] Faaß and Eckart (2013)

Definition of utterances

- Utterances were defined with the help of the pause tag in KEC.
- Its mean and median duration in the dataset were 2.018 and 0.800 seconds.

PCA on the phonological variables

- `uDur`, `uNumSyl`, `uSpRate`, `wDur`, `wNumSyl`, and `wSpRate` were correlated.

↓

- `uDur`, `uNumSyl`, `wDur`, and `wNumSyl` were combined by PCA.
  - `uSpRate` and `wSpRate` were excluded because they were calculated from `uDur`, `uNumSyl`, `wDur`, and `wNumSyl`.
- PC1 explained 99.3% of the variance.
  - → Only PC1 was included as the composite measure of speech rate.
- PC1 was correlated with...
  - `uDur` positively
  - `uNumSyl` positively
  - `wDur` negatively
  - `wNumSyl` positively

Duration model

`sDur` $\sim$ `Suffix` + `Morph` + `Suffix:Morph`

- $\rightarrow$ Effects of *-er* vs. *-t*
- $\rightarrow$ Effects of morphological status of the word-final *-er/-t*.

\+ `s(PC1, k=3)` + `uInitial` + `uFinal`

- $\rightarrow$ Phonological variables.
- $\rightarrow$ "`s()`" represents a non-linear effect.
- $\rightarrow$ "`k=3`" represents quadratic.

\+ `s(Freq, k=3)` + `s(Speaker, bs=''re'')`

- Predictability effects (`Freq`).
- Speaker variations (`Speaker`).

Study 1: Model summary

(A. Parametric)	β	SE	t	p
Intercept	-2.316	0.008	-295.116	$<$0.001
Suffix=t	-0.453	0.004	-126.131	$<$0.001
Morph=TRUE	0.061	0.008	7.742	$<$0.001
uInitial=TRUE	0.018	0.004	4.265	$<$0.001
uFinal=TRUE	0.388	0.004	107.320	$<$0.001
Suffix=t:Morph=TRUE	-0.062	0.008	-7.341	$<$0.001
(B. Smooth)	edf	Ref.df	F	p
s(Freq)	1.952	1.998	177.471	$<$0.001
s(PC1)	1.981	1.999	51.226	$<$0.001
s(Speaker)	354.235	466.000	3.412	$<$0.001

Quasi-bootstrapping to ensure phone-boundary realiability

- Major discrepancies between manual and forced alignment are rare [9]
- Small displacements do occur.
- Quasi-bootstrapping (1000 times):
  1. Resample 10% of the data without replacement.
  2. Fit the same model.
  3. Calculates the slopes of `Morph` for *-er* and *-t*.
- Bootstrapped confidence intervals:
  - *-er* $\rightarrow$ [0.011, 0.104]
  - *-t* $\rightarrow$ [-0.023, 0.019]

[9] MacKenzie and Turton (2020)

Exclusion of measurement errors (1/2)

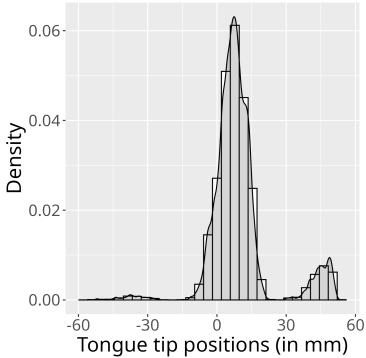
- Recordings with a sampling rate of 400 Hz [10].
  - → 0.0025 seconds for each data point.
  - → The tongue cannot move so much for 0.0025 seconds.
- Some “jumping” data points were observed.

↓

- The tokens with jumping data points were excluded.
  - Jumping data points: the data points that are more than 5 mm away from the previous data point.
- After the exclusion:
  - 138,390 data points
  - 4,678 word tokens
  - 817 word types

[10] Arnold and Tomaschek (2016)

Exclusion of measurement errors (2/2)



- The second “bump” around 50 mm above the vertical middle point.
- It turned out to be measurement errors.A

↓

- The data points smaller than -20 and larger than 20 were excluded.
- After the exclusion:
  - 123,770 data points
  - 4,196 word tokens
  - 777 word types

The numbers of word tokens and types in Study 2

	Morphemic	Non-morphemic
-er	146	956
-t	1178	1916

	Morphemic	Non-morphemic
-er	75	95
-t	377	230

Covariates and factors in Study 2

- `Freq`: Word frequency (log) from the SdeWaC corpus [11].
- `UttInitial`: Whether the word is utterance-initial.
- `UttFinal`: Whether the word is utterance-final.
- `PC1`: The composite predictor of speech rate (the same as in Study 1).
- `Speaker`: Speaker differences (as random intercept)
- `PrevSeg`: Previous segment (as random intercept)
- `NextSeg`: Next segment (as random intercept)

[11] Faaß and Eckart (2013)

Study 2: Model structure

TonguePos ~ `s(Time, k=3) + s(Time, by=Morph, k=3) + Morph`

- `s(Time, k=3)` → Tongue movements for non-morphemic *-er*/*-t*
- `s(Time, by=Morph, k=3)` → **Differences** in tongue movements for morphemic vs. non-morphemic *-er*/*-t*
- `Morph` → **Average vertical differences** between morphemic vs. non-morphemic *-er*/*-t*

+ `s(Freq, k=3) + s(PC1, k=3) + uInitial + uFinal`

+ `s(PrevSeg, bs=‘‘re’’) + s(NextSeg, bs=‘‘re’’) + s(Speaker, bs=‘‘re’’)`

Study 2: Model summary (*-er*)

(A. Parametric)	β	SE	t	p
(Intercept)	4.181	0.922	4.532	<0.001
Morph=TRUE	-0.511	0.160	-3.187	0.001
uInitial=TRUE	-0.008	0.116	-0.071	0.943
uFinal=TRUE	-0.813	0.933	-0.871	0.384
(B. Smooth)	**edf**	**Ref.df**	**F**	**p**
s(Time)	1.997	2.000	167.807	<0.001
s(Time):Morph=TRUE	1.982	2.000	35.151	<0.001
s(Freq)	1.000	1.001	2.510	0.113
s(PC1)	1.781	1.951	1.641	0.174
s(PrevSeg)	20.090	23.000	712.766	<0.001
s(NextSeg)	49.489	58.000	458.667	<0.001
s(Speaker)	31.914	33.000	1570.737	<0.001

Study 2: Model summary (*-t*)

(A. Parametric)	β	SE	t	p
(Intercept)	8.349	0.892	9.356	<0.001
Morph=TRUE	-0.103	0.053	-1.927	0.054
uInitial=TRUE	0.151	0.061	2.495	0.013
uFinal=TRUE	-0.150	0.662	-0.226	0.821
(B. Smooth)	edf	Ref.df	F	p
s(Time)	1.998	2.000	423.649	<0.001
s(Time):Morph=TRUE	1.002	1.005	0.029	0.872
s(Freq)	1.978	1.999	23.001	<0.001
s(PC1)	1.978	1.999	22.742	<0.001
s(PrevSeg)	21.196	28.000	1143.855	<0.001
s(NextSeg)	81.715	102.000	123.640	<0.001
s(Speaker)	32.957	34.000	3539.925	<0.001

Study 2: Difference curves