Supplementary Material

1 Supplementary Material 1: Experiment 1: Acoustic Analyses on Original Recordings

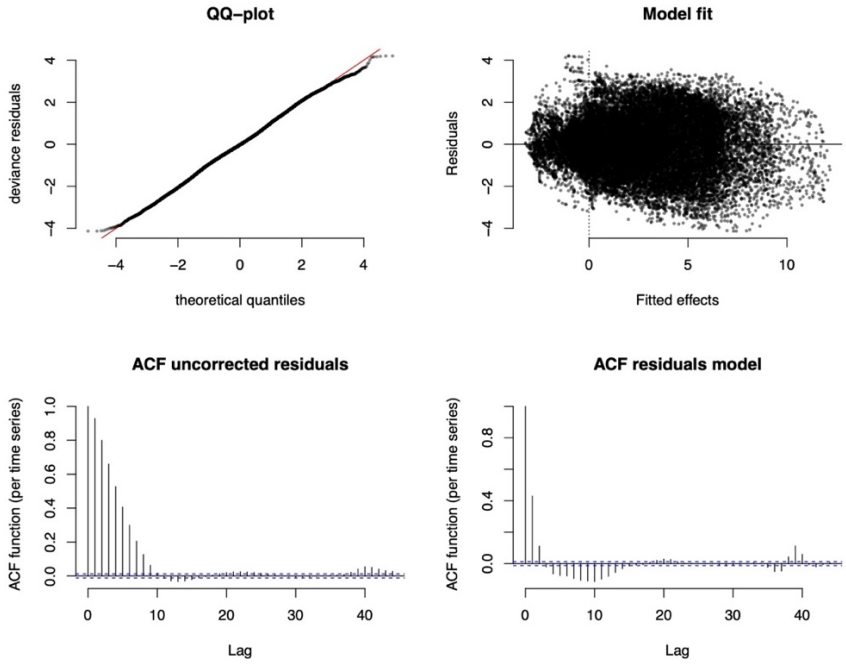
We here report on the acoustic analysis on the original recordings of the rising falling contours, which were used as a basis for PSOLA-resynthesis. Recall that recordings were done in L+H* and L*+H. Table Supplementary Material S.1.1 shows the mean values from original recordings across intonation condition (L+H* vs. L*+H) used for duration manipulation. The f0 values in the rising-falling contours were set based on the mean values in natural productions. Mean f0 values were: 166Hz for the first L (SD = 4.6 Hz), 273 Hz for the H (SD = 19.2 Hz), and 170 Hz for the second L (9.2 Hz).

Item	***wh*-element (in ms)**	**Verb (in ms)**	**Particle (in ms)**	**syllable 1 (in ms)**	**syllable 2 (in ms)**	**syllable 3 (in ms)**
Libero (verb: spielen)	123.8	340.2	207.1	170.8	146.0	334.4
Malibu (verb: trinken)	95.7	311.1	220.5	194.2	185.2	287.3
Mandalas (verb: malen)	118.0	297.2	200.7	230.6	104.7	459.6
Melanie (verb: heißen)	127.2	275.2	225.9	194.4	109.0	303.0

Supplementary Material S.1. Overview table of average durations (across original recordings in L+H* and L*+H) used for duration manipulation.

2 Supplementary Material 2: General Additive Mixed Model Results:

2.1 Evaluation of the Final Model (*model_regcond_scat_acf*).



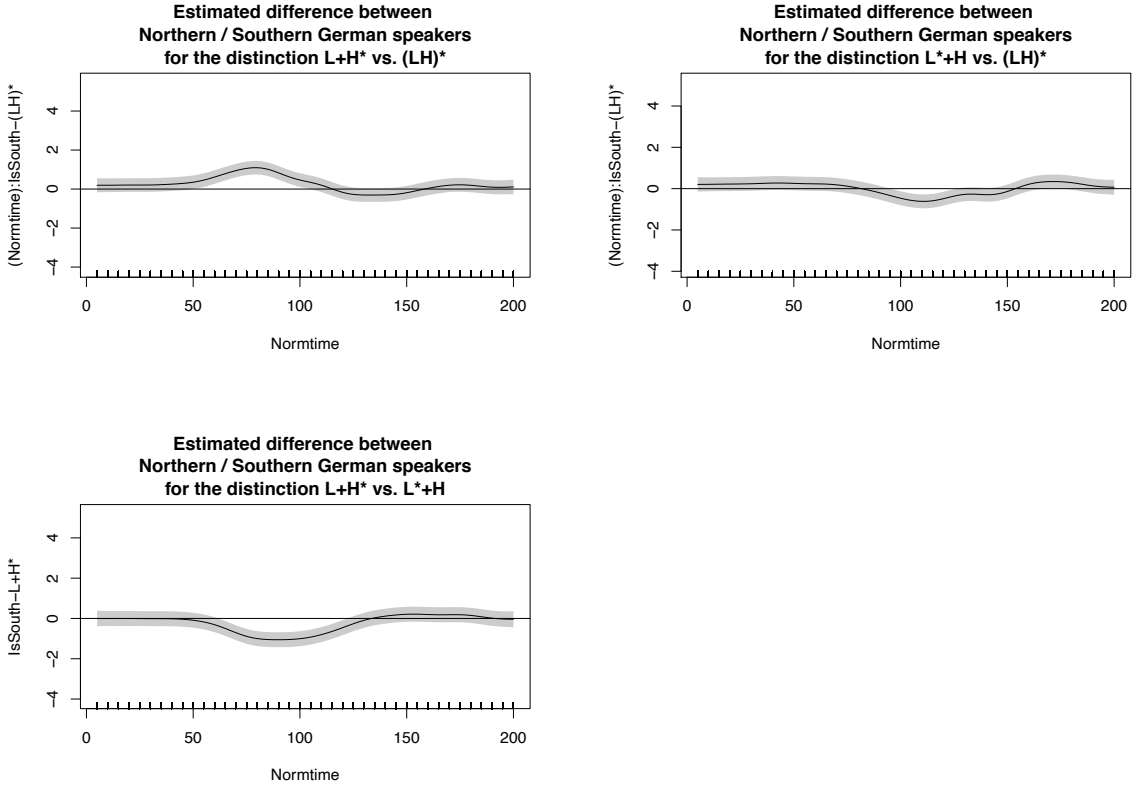
Supplementary Material S.2.1. Evaluation of the residuals in the final model, the scaled t *model_regcond_scat_acf*. Top left: Quantile-Quantile (QQ) plot (to assess potential non-normal distribution of residuals), Top right: Fitted effects plotted against residuals (to assess potential heteroscedasticity, i.e., unequal variance); Bottom left: Autocorrelation function (ACF) when residuals are not corrected; Bottom right: Autocorrelation when residuals are corrected by the rho parameter.

2.2 Interaction *Condition* x *Region* (Binary Difference Smooths)

To formally assess the interaction between intonation condition and region, we fitted models that included binary difference smooths terms that capture the difference of the difference over time between two predictors, and hence their interaction (closely following the procedure described in van Rij et al., 2019, pp. 11-13; Wieling, 2018, p. 109 ff.). To ease interpretation of these binary smooths (van Rij et al., 2019, p. 12) we fitted a GAMM model with a set of binary predictors modelling four experimental conditions (2 x 2). Hence, we assessed the effect of region in three pairwise comparisons of intonation conditions, i.e., model1: *region* (North vs. South) x *condition* (L+H* vs. (LH)*), model2: *region* (North vs. South) x *condition* (L*+H vs. (LH)*), model3: *region* (North vs. South) x *condition* (L+H* vs. L*+H)). We will explain the modelling procedure for model 1 (model 2 and 3 can be interpreted in analogy). We fitted model 1 by including a reference curve and three binary difference curves implementing the effects of *condition*, *region*, and their interaction. Specifically, we included three binary smooth terms:

- Is-(LH)* (“1” if condition is (LH)*, “0” otherwise), which implements the difference between the L+H* and (LH)* contour.
- IsSouth (“1” if region is South, “0” otherwise), which implements the difference between the Northern and Southern German speakers
- IsSouth(LH)* (“1” if condition is (LH)* and region South, “0” otherwise), which essentially implements the interaction effect that is needed to model the difference between the conditions ‘‘South.(LH)* and ‘‘South.L+H*’’ (in addition to the main effects of Is-(LH)* and IsSouth), see van Rij et al. (2019, p. 12).

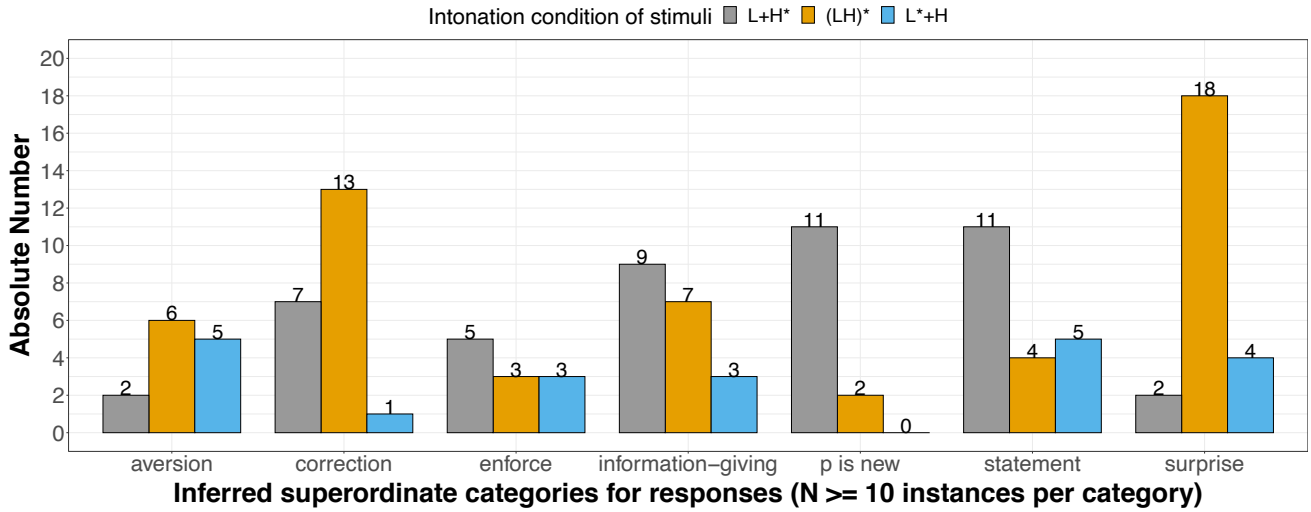
Note that models were also corrected for autocorrelation and fitted with the scaled *t* distribution specified (method = “scat”) as residuals were not normally distributed. The interaction model 1 accounted for 69.0% of the data, model 2 for 70.4%, and model 3 for 67.9%. The visualization of the the smooth term implementing the interaction effect (IsSouth(LH)* for model 1, IsSouth(LH)* for model 2, and IsSouth-L+H* for model 3) directly show when in time the distinction between contours is difference between Northern and Southern German speakers (difference of the difference), see S.2.2. Note that region affects the contour distinction in all three comparisons, but differences are small.



Supplementary Material S.2.2. Difference curve of the difference in f0 contours across region. The grey band indicates the 95% CI of the mean of the difference (across regions) of the difference (between intonation conditions). Top left: L+H* vs. (LH)* across region; Top right: L*+H vs. (LH)* across region, bottom left: L+H* vs. L*+H across region.

3 Supplementary Material 3: Post-hoc Meaning Study on Declaratives

Since Experiment 2 showed a merger of L+H* and L*+H, we tested whether for declaratives we would find a difference in the response categories for these two accents. In this follow-up study (N = 15 participants, 5 from Northern and 10 from Southern Germany), we used the same recordings of the target words (*Mandala, Malibu, Melanie, and Libero)*, spliced onto a declarative-sentence structure *Das ist der/die* 'That is the', and the same instructions and experimental procedure as in Experiment 2. Results showed differences in the interpretation between L+H* and L*+H: L+H*: was more often paraphrased as "correction", "enforcement", "statement", "p is new" and "information-giving" than L*+H. Conversely, L*+H was more often paraphrased as "surprise" and "aversion" than L+H*. In line with the results of Exp. 2, (LH)* was interpreted more often as "correction", "surprise" and "aversion" than the other two accents, see S.3.1.



Supplementary Material S.3. Distribution of response categories (inferred keywords from participants' qualitative responses, N > 10), color-coded for the different intonation conditions (L*+H in blue, L+H* in grey and (L+H)* in orange; for a split by region, the number of participants was too small.

References

van Rij, J., Hendriks, P., an Rijn1, H., Baayen, R. H., & Wood, Simon N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing, 23*, 1-22.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86-116.

Note: Dataset and analysis scripts can be found here: http://dx.doi.org/10.17632/yhv7nmjmgf.2