**Supplementary material 1**

In order to determine the optimal procedure for f0 measurements in child speech we tested the robustness and reliability of the f0 range settings:

*F0 range*

A large f0 range might increase the risk for anomalous measurements. Therefore the entire dataset was re-analysed with a smaller pitch range of 400Hz (150-650Hz). A series of correlations and kappa scores were carried out on both sets of measurements, to see if the modification of the range could have impacted the findings of the study. The results were as follows:

The Pearson’s product-moment correlation between the values reported in the present study (i.e. pitch range of 650 Hz) and the new values (i.e. pitch range of 400 Hz) was r = 0.87 (p < .001). To see whether pitch tracker settings could affect the descriptive statistics used to discuss the appearance of trochaic versus iambic pattern, the pitch distances were categorised into trochees (negative values) and iambs (positive values). The kappa coefficient between the old and new datasets was k = 0.90.

The high correlations between the reported measurements and the values obtained after narrowing the pitch range to 400 Hz, indicate that both sets of measurements yield highly similar results, which are unlikely to affect the main conclusions of the study.

*Automatic measurements*

Secondly we tested whether it is appropriate to use automatic measurements for child data without hand correction. This boils down to the question of how much noise this leaves in the data, and whether this can affect the statistical results. Therefore the pitch-tracking of all 1381 files of disyllabic utterances was visually re-inspected.

(1) 127 files of 1381 were regarded potentially problematic on the basis of visual inspection. This amounts to 9,19 % of the entire dataset. Some of the problems had to do with previously unnoticed creaky voice, but also with segmental effects of the prevocalic consonant. In these potentially problematic files, pitch what re-measured manually.

(2) The correlation between the original dataset and the new dataset with hand measured pitch values was r = 0.86 (p < .001).

(3) We also converted the semitone measurements into a nominal expression of whether pitch was higher on the first or second syllable. When comparing the original and new dataset, kappa amounted to k = 0.96.

The high correlations between the automatic and hand corrected measurements, indicate that both sets of measurements yield highly similar results, which are unlikely to affect the main conclusions of the study.