9th International Conference on Speech Prosody 2018
13-16 June 2018, Poznań, Poland

Prosodic characteristics of infant-directed speech as a function of maternal parity

Katalin Mády[1], *Uwe D. Reichel*[2], *Ádám Szalontai*[1], *Anna Kohári*[1], *Andrea Deme*[3]

[1]Research Institute for Linguistics, Hungarian Academy of Sciences
[2]Ludwig-Maximilians-Universität Munich, Germany
[3]Eötvös Loránd University, Budapest, Hungary

{mady.katalin|szalontai.adam|kohari.anna}@nytud.mta.hu,
reichelu@phonetik.uni-muenchen.de, deme.andrea@btk.elte.hu

Abstract

Infant-directed speech (IDS) has been found to be characterised by higher fundamental frequency, wider pitch range, higher energy, lower speech rate and various other factors. These tendencies occur in most languages, although with some inter-language variation. This study tries to establish these measures for the speech of Hungarian mothers of newborn babies who differ in terms of their maternal parity. Both mothers who gave birth for the first time (primipara, PP) and mothers with multiple pregnancies (multipara, MP) use higher pitch when talking to their babies as opposed to an adult. An overall increase of f0 in both speaking styles was observed for multipara mothers, although statistically not significant. IDS of this group was also characterised by higher energy and more prominent pitch accents, while these effects were missing from the IDS of primipara mothers. Directedness or parity had no effect on syllable rate.

Index Terms: infant-directed speech, directedness, maternal parity, prosodic stylisation

1. Introduction

Infant-directed speech (IDS) can be differentiated from adult-directed speech (ADS) by a number of phonetic parameters. Characteristic parameters change along with the age of the child: exaggerated prosodic features are more typical of IDS than of child-directed speech (CDS) referring to motherese used with children of 2 years or more [1]. IDS has been found to be characterised by higher fundamental frequency (f0) [2, 3, 4, 5, 6, 7], as well as wider pitch range [8, 3, 5, 7]. In most of these studies pitch was calculated using mean f0 values measured over longer speech units, such as utterances, or as in the case of [3] over a certain interval (2 minutes in their case). While there are notable exceptions [9], these differences between IDS and ADS are believed to exist cross-linguistically [5, 6, 10, 7], even if not to the same degree in all languages. [5] for example, report differences in the pitch range expansion between Japanese and German IDS. Even if IDS realisations are uniform across languages, individuals seem to show considerable variation in their IDS. [7] report on a longitudinal cross-linguistic study where they found that there was a high degree of variation in speakers' choice in the exact prosodic modification (pitch height or pitch range) used in their IDS, with some of their subjects not producing any prosodic differences between their IDS and ADS. This finding suggests that there might be a number of factors, sometimes left out of consideration, that might influence subjects' production of IDS. One of these factors might be maternal parity (primaparous, multiparous), indicating the mother's experience with IDS. While [5] did not find an effect for maternal parity in combination with infant states (drowsy, awake, etc.), [11] concludes that mothers' prosodic intensity (a measure based on subjective judgements of annotators which characterises the perceived level of infant directness of speech) decreases with each additional child with the major dividing line being between primiparous and multiparous mothers.

It has also been generally accepted that IDS has lower speech rate than ADS [3, 12, 13]. More recent evidence [14] suggests that this phenomenon is only caused by phrase-final lengthening, while local speech rate does not differ in IDS and ADS. The rhythmic properties of IDS and ADS have been barely investigated comparatively with traditional temporal measures such as duration. [15] found no deviations between IDS and ADS in terms of durational measures. However, child-directed speech (CDS), involving children above the age of 2 years, and ADS have been found to differ by several rhythm metrics (e.g. [16]). It is to be noted that these differences cannot be interpreted easily, since they are not apparent across all age groups and speech styles. Furthermore, as in the case of pitch-related phenomena in IDS, little work has been done to establish individual differences in speech rate associated with IDS and possible underlying factors.

The phonetics characteristics of IDS in Hungarian have not been widely researched. [17] conducted a study comparing mean f0 height and pitch range in ADS, IDS, and dog directed speech in a number of contexts and situations, involving both mothers and fathers. They found that IDS had higher f0 means and a larger f0 range than ADS both in spontaneous and fixed content contexts and speaker types.

The goal of this study is to check to what extent prosodic characteristics of IDS established for other languages are relevant for Hungarian caretakers. Moreover, we aim to contribute further insight to the yet understudied influence of parity on IDS, i.e. we test if mothers who give birth for the first time and thus have no history of contact with their own child start to use IDS in the same manner as mothers who have already gained substantial experience in verbal communication with their own babies.

2. Materials and methods

2.1. Participants

38 participants were recruited at the Birth Centre of the Military Hospital in Budapest. Mothers differed in terms of their parity: 22 mothers who gave birth to their first child agreed to take part in a longitudinal study up to the infant's 18^{th} month. They

10.21437/SpeechProsody.2018-60

are referred to as the primipara (PM) group. 16 mothers had at least one older child at the time of the recording. They belong to the multipara (MP) group. All participants signed an informed consent form of their recordings for scientific purposes. Data on the socioeconomic status and linguistic environment of the participants were also collected and will be included in future studies with more participants.

2.2. Materials

Recordings were based on a fairytale that was created according to the goals of phonetic analysis created by our research group. The story is about four human-like creatures with a friendly demeanour, i.e. pixies (N.B. the Hungarian term *manó* has a positive connotation). Two of them, *Tút* and *Szut*, decide to play hide-and-seek. The seeker meets two other pixies on his way, *Dát* and *Zat*, who help him to find the hider in the end. The story was presented to the participants in form of a book that contained 13 colourful paintings showing the story. On some pages, there were narrative sentences, on others, the pixies' text was shown in bubbles. Other pages just contained an image without text. Mothers were asked to familiarise themselves with the story first, then to tell the story to the research assistant (ADS), and subsequently to their own infant (IDS). Participants were instructed to tell the story with their own words, but also to read the printed sentences if present. An example is given in Figure 1.

Figure 1: *Painting from the fairytale used in the experiment. Words:* "At last, we have found you, Szut! We thought you will never turn up again!"

This design had several advantages. First, it allowed for several repetitions of the pixie names that had a CVC structure with alveolar consonants and controlled vowels. Second, certain sentence structures (i.e. prosodic boundaries) could be elicited as needed for further analysis. Third, identical sentences in both AD and ID style and across participants were available. Since the story did not make sense if only the written utterances were read aloud, participants had to create semi-spontaneous narratives when telling the story.

For the sake of comparability, only the 10 pre-formulated units were used for the present analysis. If they contained two sentences, these were regarded as two separate units. The sentence set contained 2 narrative declarative sentences. The rest was "pixie speech" including 2 yes/no questions and 13 other sentences such as vocatives, calls, and exclamatives.

Thus, the data set contained 17 sentences from 38 speakers (22 PP, 16 MP) in both ADS and IDS style. 20 sentences were either omitted by the speaker, or they differed so much from the expected form that they were excluded from further analysis. Thus, altogether 1272 utterances were analysed.

Recordings, all one or two days after birth were made in a quiet room of the hospital without any sound treatment via a Zoom H4n external sound recorder. In order to filter out potential noise from the surroundings, a head-mounted hypercardioid microphone Beyerdynamic TG H74c was attached to the recording device.

2.3. Prosody parameterisation

Next to the extraction of general f0 and energy features, we carried out a computational parameterisation of the f0 contour in the superpositional CoPaSul stylisation framework [18]. As illustrated in Figure 2, f0 is decomposed into global components here corresponding to the target sentence segments, and into local components corresponding to the vowels of each pixie name. The features will be introduced in the subsequent sections.

2.3.1. Data preprocessing

F0 was extracted by autocorrelation (Praat 6.0 [19], sample rate 100 Hz). Voiceless utterance parts and f0 outliers were bridged by linear interpolation. The contour was then smoothed by Savitzky-Golay filtering [20] using third order polynomials in 5 sample windows and transformed to semitones relative to the base value 1 Hz. Energy in terms of root mean squared deviation was calculated with a sample rate of 100 Hz in Hamming windows of 50 ms length. Syllables were extracted automatically based on the energy contour after bandpass-filtering as described in [18]. To localize the target vowels in the pixies' names, signal and text were aligned by means of the WEBMAUS webservice [21, 22].

ID and AD speech were compared based on parameters that had been shown to be prosodic correlates of hyperarticulated speech [23] and of IDS. On the sentence level, higher f0 and larger f0 range are expected in IDS, along with lower syllable rate. The increase of vowel-level prominence in pixie names can be expressed by higher energy, by a more pronounced f0 curvature, as well as by a stronger deviation of the local pitch register from the sentence-level register.

2.3.2. Sentence-level features

On the sentence level we extracted for both the f0 and energy contour the median, the maximum and the interquartile range (features $en|f0\_med|max|iqr$). Furthermore, we calculated the rate of the extracted syllables (syl_rate). In order to capture sentence-level register we fitted a base-, mid- and topline through the $[0\ 1]$-time normalized f0 contour as shown in Figure 2. Following [24] we extracted two aspects of register, namely level and range. Level is represented by the intercept, the slope, and the mean of the fitted midline ($ml\_c0|c1|m$). For range we fitted an additional regression line through the pointwise distances between base- and topline and analogously collected the intercept, the slope, and the mean of this line ($rng\_c0|c1|m$); a negative range slope value indicates that base- and topline converge as in Figure 2, a positive value indicates line divergence.

2.3.3. Vowel-level features

All vowel-level features were extracted within an analysis window of 300 ms length centered on the target vowel's midpoint. As for the sentence level we extracted the f0 and energy median, the maximum and the interquartile range. In order to make these values comparable across different positions in the sen-

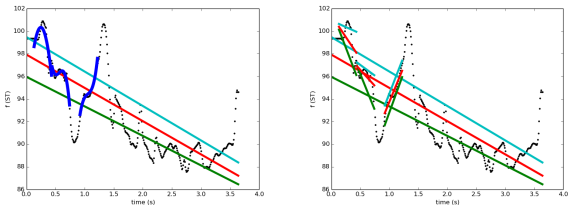

Figure 2: *Superpositional intonation stylisation of the sentence "Tút, Dát és Zat hanyatt-homlok menekült az oroszlán elől (Tút, Dát and Zat harum-scarumly fled from the lion) with three underlined target vowels on the fleeing pixies' names. Sentence-level register is captured by a base, mid, and a topline. A range regression line is fitted to the pointwise distance between the base- and the topline.* **Left:** *F0 shapes around the target vowel are represented by third-order polynomials.* **Right:** *Vowel-related f0 register is calculated analogously to the sentence level. The deviation between vowel- and sentence-related lines quantifies how much the local f0 contour sticks out from the underlying sentence-level contour.*

Table 1: *Examined f0 and energy features for the domains* sentence *(calculated over each target sentence) and* vowel *(calculated in an analysis window of 300 ms length centered on the vowel of each pixie name; normalisation window of 600 ms length and with the same midpoint).*

Feature	Description	Domain
ml_c0	f0 midline intercept	sentence
ml_c1	f0 midline slope	sentence
ml_m	f0 midline mean	sentence
rng_c0	f0 range intercept	sentence
rng_c1	f0 range slope	sentence
rng_m	f0 range mean	sentence
en_iqr	interquartile energy range	sentence
en_med	energy median	sentence
en_max	energy maximum	sentence
f0_iqr	interquartile f0 range	sentence
f0_med	f0 median	sentence
f0_max	f0 maximum	sentence
syl_rate	syllable rate	sentence
c0–3	polynomial coefs for f0 shape	vowel
ml_rms	f0 midline deviation from sentence level	vowel
rng_rms	f0 range deviation from sentence range	vowel
en_iqr_nrm	normalized interquartile energy range	vowel
en_med_nrm	normalized energy median	vowel
en_max_nrm	normalized energy maximum	vowel
f0_iqr_nrm	normalized interquartile f0 range	vowel
f0_med_nrm	normalized f0 median	vowel
f0_max_nrm	normalized f0 maximum	vowel

tence they were normalized by dividing them by corresponding reference values accounting for the local utterance context. These reference values were calculated within a longer window of 600 ms length again centered on the vowel midpoint (features *en|f0_med|max|iqr_nrm*). As shown in the left part of Figure 2 the local f0 shape around the target vowels is represented by third-order polynomials that were fitted on the $[-1\ 1]$-time normalized f0 contour after subtraction of the sentence-level register midline (features *c0–3*). In doing this *c0* captures the f0 offset in the vowel midpoint from the underlying sentence-level register, *c1* and *c3* are related to the local f0 trend (falling or rising) and to peak alignment, while *c2* represents the f0 curvature (convex or concave) and its acuity. Finally, as illustrated in the right part of Figure 2, we extracted a local vowel-related register representation in the same way as for the sentence level (cf. section 2.3.2) and quantified the amount by which the local register deviates from the sentence level register. The deviation was measured in terms of the root mean squared deviation between the vowel related mid- and range line on one hand and the corresponding stretches of the sentence related mid- and range line (features *ml|rng_rms*).

3. Results

Statistic analysis in R was based on linear mixed-effect models with directedness and parity as fixed effects and speaker and sentence as random effects. Analysis was based on random intercept models first. Subsequently, data were re-analysed with random slope models that are more conservative, in order to recheck the outcome. Since the calculation of p-values is problematic for mixed-effect models due to the absence of degrees of freedom, we adapted the thumb rule suggested by [25]: differences where $|t| > 2$ were regarded as significant.

3.1. Sentence-level features

Since pitch and energy were found to be higher in IDS in a large variety of languages, we hypothesised that all parameters listed in Table 1 regarding f0 and energy on the sentence level will have higher values in IDS compared to ADS.

IDS was indeed characterised by higher values for the parameters f0_med (t=3.91), f0_max (t=3.99), ml_c0 (t=4.31), and ml_m (t=4.69). Interestingly, f0 range measures did not differ significantly between AD and ID speech (t=−0.36). Although maternal parity and the interaction of directedness and parity did not have a main effect on the results, multipara mothers had higher f0 values for both their AD and ID fairytales for all four parameters. Figure 3 is representative for the above-mentioned measures in terms of the tendencies that both IDS style and MP status lead to higher f0 values.

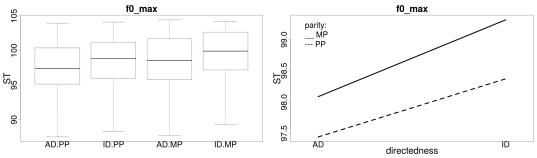

Figure 3: *F0 maximum on the sentence level. AD: adult-directed speech, ID: infant-directed speech, MP: multipara mothers, PP: primipara mothers.*

This might be an accidental finding given that speaker samples are relatively small (22 PP vs. 16 MP speakers), but it is also possible that mothers who are experienced IDS users automatically switch to a style closer to IDS when telling a fairytale, regardless whether the audience are adults or infants. However, this assumption must remain speculative until analysis can be extended to larger speaker groups.

Surprisingly, directedness did not have a main effect on energy features. This might be an artefact on the technical specifications of the hypercardioid microphone used for the recordings. However, the interaction of directedness and parity was significant for en_med (t=4.24), en_max (t=2.51), en_iqr (t=3.89) and en_rms (t=4.26). Again, Figure 4 is an example for the tendency found for all energy features mentioned above.

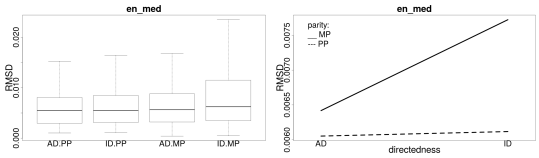

Figure 4: *Energy median on the sentence level. AD: adult-directed speech, ID: infant-directed speech, MP: multipara mothers, PP: primipara mothers.*

The same data were analysed by random slope mixed-effect models, in which a larger variation in the random effects might result in lower t-values and thus in the absence of significant differences. While directedness still has a main effect on the f0 measures (f0_med: t=2.60, f0_max: t=2.96, ml_c0: t=3.24, ml_m: t=3.07), interaction between directedness and parity becomes non-significant for the energy parameters apart from en_rms (t=2.11). This is due to high slope variation among speakers in the multipara group that is ignored in random intercept models. While the difference between the outcome of the two models might be puzzling at the first glance, we hope to be able to utilise the difference in order to detect different strategies used by mothers in their ID speech on a larger dataset.

Syllable rate did not show an impact of directedness or maternal parity. This is not surprising, since the effect of directedness was also missing in previous studies based on ADS vs. IDS comparisons, as was reported in the Introduction. Figure 5 shows an example for syllable rate.

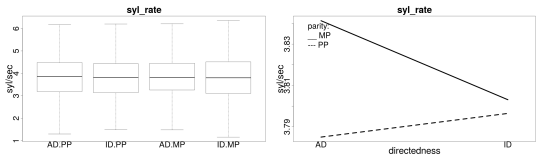

Figure 5: *Syllable rate on the sentence level. AD: adult-directed speech, ID: infant-directed speech, MP: multipara mothers, PP: primipara mothers.*

3.2. Vowel-level features

Vowel-level features were calculated for the pixie names that had comparable phonotactics, i.e. monosyllabic CVC sequences, and they predominantly occurred in comparable prosodic positions, e.g. they always carried a pitch accent. Despite of this, only two out of 12 measures showed an impact of directedness, whereas maternal parity was not reflected by any of the parameters.

The coefficients c1 and c3 of the f0 shape around the syllable centre showed that multipara mothers produce pixie names with a higher slope of the f0 (random slope models: t=2.20 and t=−3.19). They apply this strategy to the names in both speaking styles, thus, it is not dependent on directedness, at least in the present data.

3.3. Sentence types and pragmatic functions

The sentence set contained utterances that differed with respect to sentence type and function within the fairy tale. Directedness and parity were compared in the following subsets of the sentences: (1) narrative: declaratives, (2) dialogues between pixies: (2) declaratives, (3) yes/no questions, (4) exclamatives. The detailed analysis did not reveal stronger effects of directedness or parity. On the contrary, due to the data reduction resulting from the smaller number of test items, most effects became non-significant.

4. Discussion and conclusions

IDS has been investigated in a wide range of studies in many languages. However, the settings for recordings differ to a high extent, e.g. with respect to the age of the infants, speech material, the length of utterances, if speakers spoke to their own child, or if a baby was present at the recording session at all.

In this experiment, we could exclude various factors for this variation. All mothers spoke to their own babies who were one or two days old, and they all produced the same utterances. They were not asked to use baby talk, thus, they communicated in a close-to-natural way they would use with their baby – apart from the fact that some sentences were fixed.

Higher f0 values reported for ids in many languages were also found in our data. However, f0 range did not turn out to be higher in motherese speech. Interestingly, the utilisation of higher energy in IDS was only characteristic for the multipara group, while primipara mothers did not use higher energy when talking to their infants. Again, vowels in the pixie names were more prominent in the multipara group, at least with respect to some of the features investigated. This effect did not show an influence of directedness.

A surprising outcome of the study was that some of previously reported findings associated with IDS could not be replicated. We did not find that pitch range was affected by directedness, and speech rate in our data was not slower for IDS as compared to ADS. It is to be tested whether this is an artefact of the speech material, i.e. if the fairy tale triggered a certain speaking style even in ADS. This question will be answered by future ADS data collection from free dialogues with adults.

In this study, we were able to gain data from primipara and multipara mothers with newborn babies of the same age. This allowed us to ask the question whether IDS is influenced by previous experience with other children. While in most cases, both primi- and multipara mothers used similar strategies in our samples, the multipara group did show more "IDS-like" features: they used higher f0 even in their AD speech, their samples had higher energy in IDS, and their pitch accents showed more emphasis. The results suggest that the intensity of IDS increases and the audience for IDS broadens with higher experience with infants. Since the present data is part of an ongoing study, future analysis of an extended data set will shed more light on potential other influential factors of parity on IDS speech.

5. Acknowledgements

This research was funded by the NKFI grant nr. 115385 (PI István Winkler, Institute of Cognitive Neuroscience and Psychology, HAS). The work of the second author was financed by a grant of the Alexander von Humboldt Foundation. In addition, we would like to thank Beáta Gyuris for her valuable comments on the classification of the sentence categories.

6. References

[1] S. Vosoughi and D. K. Roy, "A longitudinal study of prosodic exaggeration in child-directed speech," in *Proceedings of the 6th International Conference on Speech Prosody*, 2012, p. 194197.

[2] C. A. Ferguson, "Baby talk in six languages," *American Anthropologist*, vol. 66, no. 6, pp. 103–114, 1964.

[3] A. Fernald and T. Simon, "Expanded intonation contours in mothers' speech to newborns," *Developmental Psychology*, vol. 20, no. 1, pp. 104–113, 01 1984.

[4] A. Fernald and P. Kuhl, "Acoustic determinants of infant preference for motherese speech," *Infant Behavior and Development*, vol. 10, no. 3, pp. 279–293, 1987.

[5] A. Fernald, T. Taeschner, J. Dunn, M. Papousek, B. de Boysson-Bardies, and I. Fukui, "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," *J Child Lang*, vol. 16, no. 3, pp. 477–501, 1989.

[6] D. L. Grieser and P. Kuhl, "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese," *Developmental Psychology*, vol. 24, pp. 14–20, 01 1988.

[7] C. R. Narayan and L. C. McDermott, "Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations," *J. Acoust. Soc. Am.*, vol. 139, no. 3, pp. 1272–1281, Mar 2016.

[8] D. N. Stern, S. Spieker, R. K. Barnett, and K. MacKain, "The prosody of maternal speech: infant age and context related changes," *J Child Lang*, vol. 10, no. 1, pp. 1–15, Feb 1983.

[9] B. N. Rattner and C. Pye, "Higher pitch in BT is not universal: acoustic evidence from Quiche Mayan," *J Child Lang*, vol. 11, no. 3, pp. 515–522, Oct 1984.

[10] C. Kitamura, C. Thanavishuth, D. Burnham, and S. Luksaneeyanawin, "Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language," *Infant Behavior and Development*, vol. 24, pp. 372–392, 2002.

[11] M. Monnot, "Function of infant-directed speech," *Human Nature*, vol. 10, no. 4, pp. 415–443, 1999.

[12] R. P. Cooper and R. N. Aslin, "Preference for infant-directed speech in the first month after birth," *Child Dev*, vol. 61, no. 5, pp. 1584–1595, 1990.

[13] J. S. Tang and M. J. A., "Prosodic aspects of child-directed speech in Cantonese," *Speech Hear. Lang. 9*, vol. 9, pp. 257–276, 1996.

[14] A. Martin, Y. Igarashi, N. Jincho, and R. Mazuka, "Utterances in infant-directed speech are shorter, not slower," *Cognition*, vol. 156, pp. 52–59, 11 2016.

[15] C. Lee, C. Kitamura, D. Burnham, and N. P. McAngus Todd, "On the rhythm of infant- versus adult-directed speech in Australian English," *J. Acoust. Soc. Am.*, vol. 136, p. 357, 07 2014.

[16] E. Payne, B. Post, P. Prieto, L. Astruc, and M. Vanrell Bosch, "Rhythmic modification of child directed speech," *Oxford University Working Papers in Linguistics, Philology and Phonetics*, vol. 12, pp. 147–184, 2010.

[17] A. Gergely, T. Faragó, A. Galambos, and J. Topál, "Differential effects of speech situations on mothers' and fathers' infant-directed and dog-directed speech: An acoustic analysis," *Sci Rep*, vol. 7, no. 1, p. 13739, 2017.

[18] U. Reichel, *CoPaSul Manual – Contour-based parametric and superpositional intonation stylization*, RIL, MTA, Budapest, Hungary, 2017, https://arxiv.org/abs/1612.04765.

[19] P. Boersma and D. Weenink, "PRAAT, a system for doing phonetics by computer," Institute of Phonetic Sciences of the University of Amsterdam, Tech. Rep., 1999, 132–182.

[20] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, no. 8, pp. 1627–1639, 1964.

[21] F. Schiel, "Automatic Phonetic Transcription of Non-Prompted S peech," in *Proc. ICPhS*, San Francisco, 1999, pp. 607–610.

[22] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer, Speech, and Language*, vol. 45, no. C, 2017.

[23] B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, W. Hardcastle and A. Marchal, Eds. Dordrecht: Kluwer, 1990, pp. 403–439.

[24] T. Rietveld and P. Vermillion, "Cues for Perceived Pitch Register," *Phonetica*, vol. 60, pp. 261–272, 2003.

[25] R. H. Baayen, *Analyzing linguistic data*. Cambridge: University Press, 2008.