(under)graduate journal
LingUU
linguistics

BEYOND BABBLING: THE USE OF PROSODY IN VOCALIZATIONS OF MONOLINGUAL DUTCH AND SWEDISH 3-MONTH-OLD INFANTS

A CROSS ACOUSTIC ANALYSIS OF PITCH AND COMMUNICATIVE FUNCTION IN EARLY INFANCY

Julia Heemskerk, Patricia Dworak, Yasmina Forsch: *Julia.heemskerk@hotmail.com; patiewelina.71@gmail.com; yasminaforsch@gmail.com*
Institute for Language Sciences, Utrecht University, Utrecht, the Netherlands

KEYWORDS
Infant prosody
Pitch modulation
Communicative intent
Cross-linguistic development
Early vocalizations

INFO
Received: January 11, 2026 ■
Accepted: May 11, 2026 ■

ABSTRACT
This exploratory study examines whether 3-month-old infants use pitch to signal communicative intent (requests vs. comments) and whether such prosodic modulations differ across languages. An analysis was conducted on spontaneous vocalizations (natural speech) from monolingual Dutch and Swedish infants (N = 9), extracted from an existing dataset and study by van Niekerk, Junge & Chen (2025). Infant vocalizations were annotated as 'requests' or 'comments' and analyzed for two acoustic features: mean pitch (average pitch) and pitch span (the spread between high and low pitch). Linear mixed-effects models revealed that 'requests' were produced with higher mean pitch and wider pitch span than 'comments' across both language groups. No main effect of language was found for mean pitch, whereas language difference emerged for pitch span, indicating that the distinction between requests and comments was more pronounced in Dutch than in Swedish infants. These findings suggest that intentional prosodic signalling is present by 3 months of age, and early signs of language-specific prosodic development may already be observed in pitch span.

1. INTRODUCTION

From early coos and cries, infants demonstrate sensitivity to the acoustic properties of speech long before the emergence of words. Research on early vocal development has shown that infant vocalizations provide insight into the foundations of communication, with prosody playing a central role in early language acquisition (Oller, 2000; Jhang & Oller, 2017). Among prosodic features, pitch has been identified as a particularly salient and socially meaningful cue, shaping early interaction and signaling communicative intent (Fernald, 1993; Trainor et al., 2000).

The present study investigates how 3-month-old infants begin to use pitch in their vocalizations, focusing on its potential role as a marker of communicative intent. This age is particularly significant as infants exhibit volitional control over vocalization and respond to social cues with distinct prosodic patterns (Buder et al., 2013), yet precedes canonical babbling and perceptual narrowing, allowing capture of vocal modulation in its earliest intentional form (Werker & Tees, 1984). A second aim is to investigate whether early prosodic

differentiation reflects universal developmental patterns or is influenced by language-specific input. A recent systematic review by van Niekerk, Junge, & Chen (2025) highlighted that cross-linguistic studies on early pitch production remain scarce, especially for the 0–3-month range. This exploratory study compares Dutch and Swedish infants because of fundamental differences in their prosodic systems (Bruce, 2007; Chen, 2011; Romøren, 2016), providing a valuable test of whether early prosodic production is shaped by language input or reflects universal patterns of development. The current study analyzes a subset of data from a larger dataset collected by van Niekerk, E., Junge, C., Schwarz (2025). This dataset was selected as part of a research project at Utrecht University.

2. THEORETICAL FRAMEWORK

2.1. Prenatal and early sensitivity to prosody

To interpret pitch modulation in 3-month-old infants, it is essential to first examine prenatal and early sensitivity to prosody. Before birth, fetuses are exposed to rhythmic and melodic features of speech through amniotic fluid, including pitch contours and sentence intonation (Gervain et al, 2021). This early auditory exposure likely forms the foundation for prosodic learning, which is hypothesized to support early intention signaling through pitch variation (Mampe et al., 2009). Evidence for this early sensitivity comes from Mampe et al. (2009), who showed that newborns produce cry melodies that reflect the intonation patterns of their native language, such as rising contours in French and falling contours in German. These cross-linguistic differences suggest that prenatal exposure influences the prosodic structure of early vocalizations. In addition, Martínez-Álvarez et al. (2023) found that newborns can perceptually distinguish between well-formed and deviant prosodic contours, as shown by increased right-hemisphere activation. Together, these findings indicate that prosodic sensitivity is present from birth and may already support early communicative signaling.

In line with this, van Niekerk, Junge, & Chen (2025) propose that infants' early vocalizations, particularly protophones, can serve as functional roles even before they become linguistic. Protophones, as defined in Buder et al. (2013), are voluntarily produced sounds with an organized acoustic structure, such as vocants, squeals, and growls. The study of van Niekerk, Junge, & Chen (2025) argues that preverbal infants may use vocal signals to influence caregivers and initiate interaction. This indicates that protophones are not merely shaped by social exposure but reflect emerging communicative intent. Buder et al. (2013) likewise emphasize the volitional nature of protophones and position them as precursors to intentional prosody. These insights align with the current study's focus on whether infants modulate pitch intentionally in response to communicative function.

2.2. Early communicative intentions in pitch

The development of vocalizations in early infancy suggests that young infants exhibit a degree of volitional control over their vocal apparatus (Buder et al., 2013). In this notion of early control, a distinction can be made between reflexive sounds (such as crying or burping) and protophones. These early sounds demonstrate emerging control over vocal folds and articulators and form the basis for melodic variation and intentional prosodic modulation. In line with this, Wermke et al. (2021) observed a gradual increase in melodic complexity in both cry and vocal utterances from two to three months of age. Non-cry vocalizations from that age onwards show multiple pitch movements within a single utterance, indicating an early modulatory capacity. Prosody is thus not merely an affective side-effect but it is actively regulated in social and communicative contexts. This early control over pitch and vocal modulation forms the basis for what is referred to in this study as intentional prosodic signaling: systematic variation in pitch according to communicative function. In this study, 'comments' are defined as vocalizations produced without the intention of requesting something, often serving a descriptive or affective function (Esteve-Gibert & Prieto, 2013; Buder et al., 2013). Theoretical accounts suggest that communicative intent in pitch variation, particularly higher mean pitch

and wider pitch span, is associated with urgency, emphasis, or the intention to request, whereas lower pitch and narrower span are associated with more neutral or commentary functions (Papaeliou & Trevarthen, 2006; Esteve-Gibert & Prieto, 2013). These differ from 'requests', which aim to elicit a response from the caregiver. This functional distinction forms the basis for analyzing how infants use pitch to express communicative intent.

2.3. Cross-linguistic variation in prosodic input

The current study focuses on the comparison between Dutch and Swedish, which offers a unique opportunity to investigate how different prosodic environments shape pitch-based communication. Although both languages use pitch for communicative purposes, the languages differ significantly in the way pitch functions within their respective prosodic systems, potentially creating different pitch environments for growing infants. As both Dutch and Swedish are non-tonal, Indo-European languages, this allows for more controlled comparison. Dutch employs a stress-based prosodic system in which pitch changes primarily mark sentence-level pragmatic contrasts such as focus and interrogative mood (Haan, 2002; Chen, 2011). Questions in Dutch are typically characterized by rising pitch contours, especially toward utterance ends (Haan, 2002), creating a straightforward pitch environment where pitch reflects information structure without affecting lexical identity. In contrast, Swedish uses a lexical pitch system in which pitch contours distinguish word meanings through two pitch accents: Accent 1 and Accent 2 (Bruce, 2007; Riad, 2014). Swedish employs falling pitch contours in both statements and questions, meaning pitch change in requesting contexts does not rely on rising patterns (Gosselke Berthelsen, 2021). This results in a more complex pitch environment requiring integration of tonal targets from multiple domains. Consequently, Swedish-learning infants are exposed to both word-level and sentence-level pitch variation, whereas Dutch-learning infants primarily experience pitch as a cue to pragmatic or syntactic function.

Early language development is considered to be largely universal across languages, particularly in the prelinguistic stage, where infants' productions show limited and inconsistent evidence of language-specific differentiation. The prosodic differences between Dutch and Swedish therefore create contrasting developmental environments that allow us to test whether early pitch-based communication remains universal or begins to diverge by 3 months of age. Research by Romøren (2016) suggests that Swedish children were more consistent in pitch-based focus markers than Dutch peers, indicating that early exposure to lexical pitch distinctions may influence developmental trajectories. However, during early infancy, children remain in a stage of language-universal perception, broadly sensitive to prosodic contrasts regardless of native language structure (Werker & Tees, 1984). This suggests that infants may initially rely on universal communicative biases rather than language-specific prosodic mappings (Esteve-Gibert & Prieto, 2013; Papaeliou & Trevarthen, 2006), pointing to a developmental stage in which infants systematically use pitch to signal intent prior to the emergence of language-specific constraints.

2.4. Perceptual narrowing, prosodic bootstrapping and universality

To better understand whether early prosodic behavior is shaped by universal mechanisms or language-specific input, it is essential to consider the concept of perceptual narrowing. Perceptual narrowing states that young infants initially have a universal auditory discrimination ability, which allows them to distinguish sounds from different languages (Werker & Tees, 1984). As they are exposed to language-specific input, this ability narrows. Research showed that infants are still able to discriminate sound contrasts from unfamiliar languages between the ages of six and eight months, but that this ability largely disappears by the age of twelve months (Werker & Tees, 1984).

Although perceptual narrowing was first studied in relation to speech sounds, research shows that it also applies to broader features of speech, such as prosody. Mampe et al. (2009) showed that newborn cry melodies already display language-specific properties, presumably because of prenatal exposure to the prosodic contours of the native language. This suggests that elements such as pitch can be influenced

by environmental input even before birth. At the same time, this early sensitivity to prosody implies that similarities in pitch use by, for example, three-month-old infants, may indicate universal patterns in vocal development rather than differentiated input effects. Complementing this framework, the prosodic bootstrapping hypothesis, infants can use prosodic features such as pitch, duration, and rhythm to access abstract grammatical or lexical structure posits that infants use prosodic cues available in their input to learn the abstract lexical and grammatical properties with which these cues are correlated (Gervain et al., 2021). According to Gervain et al. (2021), experimental studies suggest that infants can indeed exploit such correlations to acquire the lexicon and grammar of their native language.

By studying infants' behavior before the so-called reorganization phase, in which perceptual narrowing actively occurs, one can thus better distinguish between innate sensitivity to prosodic features and later language-specific adaptation (Mampe et al., 2009). Comparing Swedish and Dutch 3-month-olds offers an opportunity to examine whether prosodic modulation is universally driven or already shaped by language-specific input. Consequently, if infants from different language backgrounds show similar prosodic modulation at 3 months, this would support the claim that early prosodic intent is part of a universal communicative toolkit, not yet shaped by language-specific input.

3. HYPOTHESIS

The reviewed literature indicates that infants are sensitive to prosodic structure from birth and gradually gain volitional control over pitch. While studies on older infants have demonstrated that prosody can express communicative intent (Papaeliou & Trevarthen, 2006; Esteve-Gibert & Prieto, 2013), emerging evidence suggests that the foundations of this ability may already be present at 3 months (Buder et al., 2013; Wermke et al., 2021). However, few studies have targeted prelinguistic vocalizations in early infancy, with existing research on communicative intent and prosody primarily focusing on infants aged 7 months and older. Furthermore, cross-linguistic research exploring prosodic use and communicative intent at 3 months is absent, representing a significant gap in understanding how fundamental pitch-intent relations develop across linguistic environments.

The current study addresses this gap by investigating whether pitch modulation distinguishes communicative functions at 3 months of age, a particularly significant developmental window representing a major transition from reflexive to volitional pitch control (Buder et al., 2013; Jhang & Oller, 2017). By comparing Dutch and Swedish infants, the study explores whether early pitch signaling reflects a universal communicative capacity or is shaped by language-specific prosodic input. This leads to the central research question:

How do Dutch and Swedish 3-month-old infants use pitch to distinguish between requests and comments in mother-infant interactions?

Based on previous research, the hypothesis is that infants at 3 months will already show systematic use of pitch features to signal communicative intent. This is grounded in the prosodic bootstrapping hypothesis (Gervain et al., 2021) and findings that melodic contours in infant-directed speech convey affective and attentional intent (Fernald, 1989). Two predictions follow:

1. Infants will produce requests with higher mean pitch and wider pitch span than comments. This is based on findings that communicative vocalizations are acoustically distinct, characterized by higher pitch, more pitch variation, and more dynamic contouring (Papaeliou & Trevarthen, 2006; Esteve-Gibert & Prieto, 2013).
2. No differences are expected between Dutch and Swedish infants in mean pitch and pitch span. Based on the perceptual narrowing framework (Werker & Tees, 1984), which shows language-specific tuning begins between 6 and 12 months, cross-linguistic similarity at 3 months would indicate that prosodic intent is part of a universal communicative repertoire not yet shaped by language-specific input.

4. METHODOLOGY

4.1. Participants and data

In this study, a subset of spontaneous infant vocalizations was analyzed from the larger study collected by van Niekerk et al. and van Niekerk, Junge, & Chen (2024; 2025). This dataset includes data from infants acquiring Dutch, Swedish, or Indonesian. For the purposes of this study, only Dutch and Swedish data were included. This was mainly due to time limitations, which were tied to the exploratory nature of the study and the focus on comparing two languages that differ in prosody but are relatively similar in lexical development. To control for cross-linguistic input effects, only monolingual infants were selected, as both input frequency and structural features of the language may shape early prosodic development (Han & Kager, 2020). The dataset included a total of 10 Dutch and 8 Swedish infants, all 3 months of age. For this exploratory analysis, a smaller subset was extracted, resulting in a final sample of 9 monolingual infants (5 Dutch, 4 Swedish). Across these infants, a total of 679 usable vocalization tokens were identified. Only canonical protophones, defined as non-reflexive, volitional vocalizations such as vocants, squeals, and growls, were included, in accordance with Buder et al. (2013). Reflexive vocalizations such as cries, and vegetative sounds were excluded to ensure that only vocalizations involving intentional pitch modulation were analyzed.

4.2. Materials and procedure

The materials and procedural process in this section refers to the experimental design and process in the original study which our dataset is from. Their data for this study were obtained through semi-structured mother-infant (mi) play sessions of approximately 15 minutes, recorded in the participants' homes. This type of set-up allows for the infant to feel comfortable which evokes further vocalizations. Each infant participated twice between 3;00 and 3;30 months of age. During each session, infants were observed in four different contexts within the mother-infant sessions: playing with familiar toys (mift), playing with unfamiliar toys (miut), engaging in free interaction (mi), and being alone (i). Mothers were instructed to interact with their infants as naturally as possible, with the goal of eliciting spontaneous vocalizations, particularly those functioning as requests or comments. The dataset was provided in a pre-annotated format, and therefore the selection and initial segmentation of the recordings were not performed as part of the present study. Both video and audio were recorded, the present study focused exclusively on the audio data as this data was available for analysis. In the data that was provided, vocalizations were segmented and annotated using ELAN (The Language Archive, 2024), based on an internally developed coding protocol informed by prior work on early functional vocal behavior (van Niekerk, Junge, & Chen, 2024; van Niekerk, Junge, & Schwarz, 2025). Each vocalization received three labels: (1) communicative function (request or comment), (2) interactional contexts, and (3) vocalization type (protophone or other). These annotations were adopted as provided and were not re-coded or independently validated within the scope of this study. The segmentation followed a tier-based structure consistent with the setup used in ProsodyPro, see appendix B for an example (Xu, 2013).

For this study, the files were exported and manually aligned with the audio recordings using Praat's TextGrid format (Boersma & Weenink, 2025). Then, communicative intent was annotated by three coders. Inter-rater reliability was established through discussions among the coders, and any discrepancies were resolved by consensus. Acoustic features were then extracted using the ProsodyPro script (Xu, 2013). Following established conventions for infant speech analysis, the pitch floor was set to 75 Hz and the pitch ceiling to 800 Hz. Segmentation was based on the "target tier" (tier 7) of ProsodyPro (appendix B). Two acoustic

variables were extracted for each vocalization: mean pitch (F0_mean) and pitch span (F0_max – F0_min), which were the dependent variables of the research, both measured in Hertz.

4.3. Statistical analysis

To test the hypotheses, linear mixed-effects models were applied using the lmer function from the lme4 package in RStudio (Bates, et al., 2015; Posit team, 2025). Mixed models were chosen because they account for the non-independence of observations within individuals (infants), which is critical when using repeated measures. Random slopes were not included due to the limited number of vocalizations per infant and convergence issues in model estimation. Three fixed effects were included in the model: (1) communicative function (request vs. comment), (2) language (Dutch vs. Swedish), and (3) the interaction between function and language. A random intercept for infants was added to account for individual variation in pitch. Infants were initially coded using language-specific identifiers (e.g., "s1" for Swedish, "d1" for Dutch), reflecting their language group. These were later re-coded into anonymized labels ("a" through "i") to allow for more neutral statistical modelling and avoid confounding by language in random effects structure. The interactional context was not included as a random effect due to limited data points per session and convergence issues in preliminary model testing. No additional covariates were entered into the model, as individual-level variability was accounted through the inclusion of a random intercept. In the coding of predictors, Swedish was coded as 1 (Dutch = 0), and request was coded as 1 (comment = 0). This dummy coding allows for straightforward interpretation of main and interaction effects.
To assess model fit, each dependent variable (mean pitch and pitch span) was analyzed in four step models. The four models are listed in the appendix. The model consisted of the dependent variable (mean pitch or pitch span), the fixed effects and the random intercept infants.

5. RESULTS

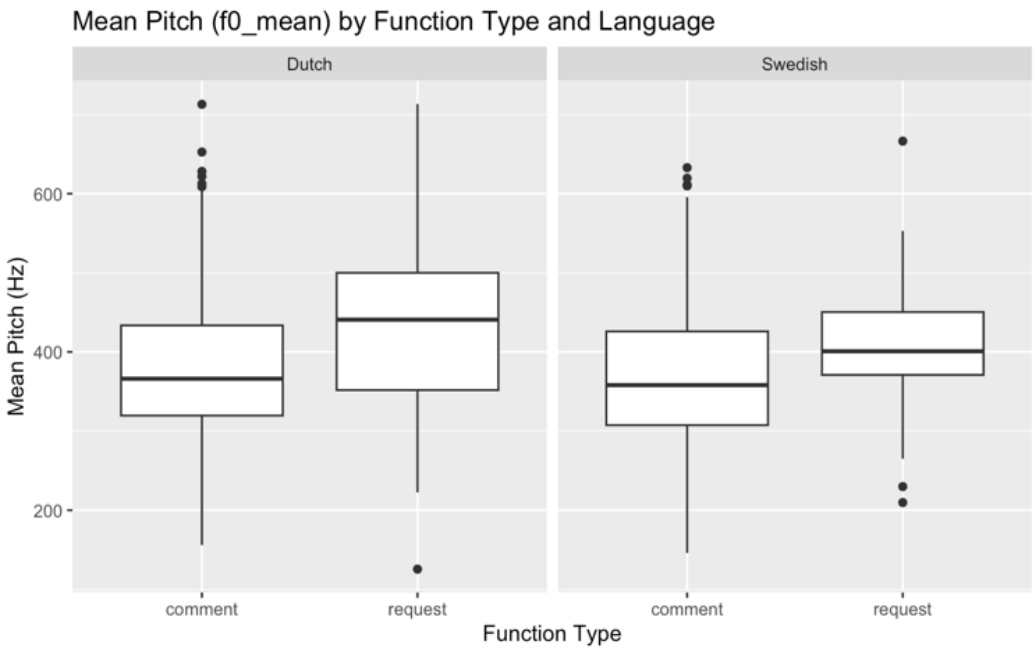
The results reveal that infants at 3 months already modulate pitch according to communicative function, with requests showing higher mean pitch and wider pitch span than comments. Language effects were minimal for mean pitch, but a significant interaction emerged for pitch span, suggesting early cross-linguistic differences in prosodic modulation.

5.1. Mean pitch

The null model, Model A1, which included only random intercepts, revealed significant variability between individual infants (*Intercept: b* = 388.77, *SE* = 11.91, $p < .001$), indicating that mean pitch varies substantially across participants. When function type was added as a fixed effect, a clear effect emerged: requests were produced with significantly higher mean pitch, approximately 42 Hz higher, than comments (b = 41.86, *SE* = 7.38, $t(676) = 5.67$, $p < .001$). In contrast, including language (Dutch vs. Swedish) as an additional predictor did not improve model fit, and the main effect of language was not significant (b = –13.24, *SE* = 21.26, $t(676) = -.62$, $p = .559$). The final model included an interaction term between function type and language. Here, only the main effect of function remained significant, where request had approximately 51 Hz higher mean pitch than comments (b = 51.44, *SE* = 22.11, $t(676) = -.30$, $p < .001$). Neither the main effect of language (b = –6.53, *SE* = 14.76, $t(676) = -1.27$, $p = .778$) nor the interaction between function and language reached significance (b = –18.72, *SE* = 10.56, $t(676) = 4.87$, $p = .205$). These results suggest that the functional distinction between requests and comments, in terms of mean pitch, is consistent across both languages. This pattern is visualized in the corresponding boxplot in Figure 1, which shows that the distribution appears wider in Swedish, with several high-pitch outliers, but the general pattern of elevated pitch for requests holds across languages. These findings confirm the first prediction: infants produced requests with higher mean pitch, regardless of language. The absence of a significant interaction supports the second prediction that mean pitch modulation does not differ cross-linguistically. This may indicate that mean pitch serves

Figure 1

Boxplots mean pitch by function and language.



Note. Function type refers to communicative function (comment and request).

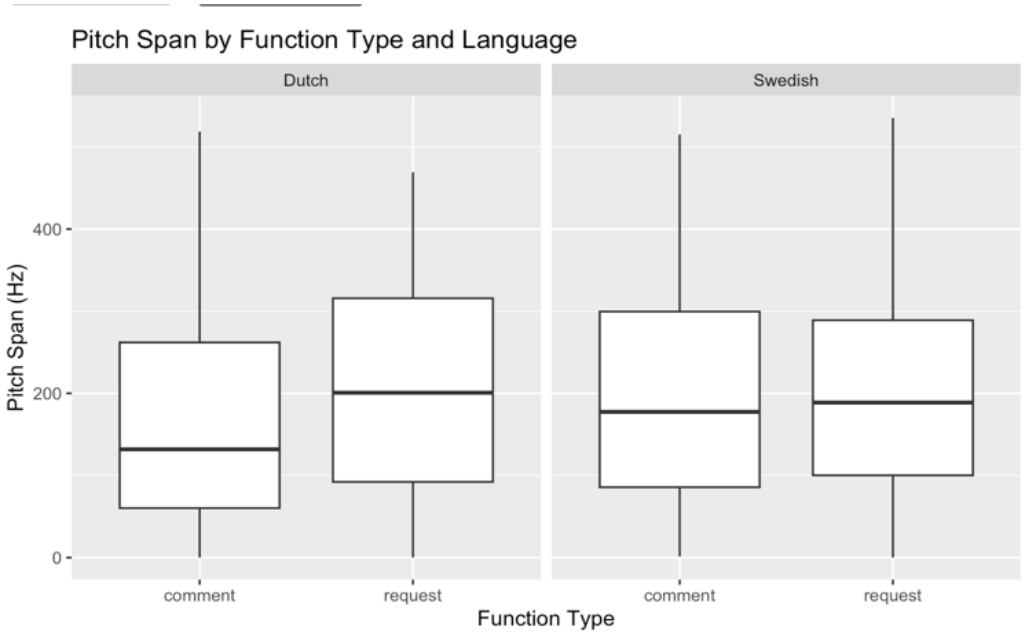
as a potentially universal prosodic cue to communicative function, even before lexical or morphosyntactic categories are in place.

5.2. Pitch span

In line with the results for mean pitch, the null model again confirmed significant individual variation in pitch span (*Intercept: b* = 187.71, *SE* = 15.45, *p* < .001). Adding function type as a fixed effect revealed that requests were produced with a significantly wider pitch span than comments, approximately with 24 Hz wider (*b* = 23.55, *SE* = 9.85, *t*(677) = 2.39, *p* = .017). Language did not emerge as a significant main effect (*b* = 25.21, *SE* = 31.16, *t*(677) = 0.81, *p* = .445). However, when including the interaction term, a different pattern emerged: function type remained a significant predictor where request had approximately 46 Hz wider pitch than comments (*b* = 46.23, *SE* = 14.05, *t*(677) = 3.29, *p* = .001), the main effect of language remained non-significant (*b* = 40.96, *SE* = 31.85, *t*(677) = 1.28, *p* = .234), but the interaction between function type and language was significant (*b* = –44.81, *SE* = 19.65, *t*(677) = -2.28, *p* = .023). The interaction showed that the differences between request and comments were approximately 44 Hz less for Swedish infants, than for Dutch infants. This interaction indicates that the contrast in pitch span between requests and comments was more pronounced in Dutch than in Swedish. This pattern is clearly visible in the boxplot in Figure 2: the functional differentiation in pitch span is much stronger for Dutch infants. This suggests that Dutch infants exhibit greater pitch range modulation for communicative functions than Swedish infants, possibly reflecting early language-specific tuning of prosodic categories. This challenges the second prediction of no language effect. While the first prediction is again supported, requests showed a wider pitch span, the second is only partially confirmed, as cross-linguistic variation emerged. This finding suggests that infants' use of pitch span may already be influenced by the specific prosodic patterns of their native language. While mean pitch appears to function as a general cue to communicative intent across languages, pitch span seems more sensitive to language-specific input.

Figure 2

Boxplots pitch span by function and language.



Note. Function type refers to communicative function (comment and request).

6. DISCUSSION

This exploratory study examined whether 3-month-old infants modulate prosodic features to signal communicative intent and whether such modulation is influenced by cross-linguistic prosodic variation. The results lend support for the main hypothesis, demonstrating that infants as early as 3-months use prosody systematically to signal communicative functions.

Specifically, the analysis revealed significant effects of function type on both mean pitch and pitch span, with requests consistently produced with higher mean pitch and wider pitch span than comments, confirming the first prediction. These findings extend previous work by Esteve-Gibert and Prieto (2013) and Papaeliou and Trevarthen (2006), which demonstrated functional pitch differences in infants aged 7 months and older, suggesting that prosodic modulation for communicative purposes may begin considerably earlier, even before canonical babbling emerges.

The second prediction received partial support. While no main effect of language was found for mean pitch, supporting the view that early prosodic production follows universal developmental trajectories (Werker & Tees, 1984), a significant interaction effect emerged for pitch span. Dutch infants showed significantly wider pitch span modulation between requests and comments compared to Swedish infants, suggesting early adaptation to language-specific prosodic input. This finding aligns partly with earlier research (Astruc et al., 2013; Mampe et al., 2009; Romøren, 2016) and indicates that while early vocalizations are initially shaped by universal mechanisms, language-specific prosodic tuning may begin earlier than previously assumed.

One possible explanation for this cross-linguistic difference lies in the prosodic structure of the input languages. Dutch is a stress-based language in which pitch primarily marks pragmatic functions such as focus and sentence type (Haan, 2002), potentially providing clearer and more consistent mappings between pitch variability and communicative intention. On the other hand, because Swedish is a pitch-accent language where pitch contours serve both lexical and pragmatic functions (Bruce, 2007; Riad, 2014), this may introduce ambiguity for infants whose prosodic control is still developing, making them less dependent on pitch span for intentional signaling at this stage. Notably, infant-directed speech (IDS) patterns may play a

role: previous work suggests that Dutch caregivers use exaggerated prosodic contours in IDS (Han & Kager, 2020; Han et al., 2022), which could enhance prosodic cues for infants. However, further research measuring caregiver input directly is needed to test these interpretations.

These findings have broader theoretical implications. They support prosodic bootstrapping frameworks (Gervain et al., 2021) and demonstrate that pragmatic competence arises through form-function mappings from early infancy. The results build on cry melody studies (Mampe et al., 2009; Wermke et al., 2021) by showing that intentional prosodic modulation also occurs in non-cry vocalizations and aligns with research on protophone functional flexibility (Jhang & Oller, 2017). Understanding these early prosodic patterns could help identify atypical prosodic development, offering potential applications for early intervention and clinical assessment of communication delays or speech disorders.

From a methodological perspective, this study has notable strengths. The use of spontaneous interaction data improves ecological validity, providing realistic illustrations of infant prosody in natural contexts. The application of mixed-effects models enabled inclusion of individual variation, and the cross-linguistic design contributes to typologically comparative research on prosodic acquisition (Bruce, 2007; Romøren, 2016). However, there are several limitations that must be acknowledged. First, the small sample size (n = 9) limits generalizability, particularly for interpreting subtle cross-linguistic differences that may require larger samples to detect reliably. Second, time constraints prevented inclusion of Indonesian infants, limiting typological comparisons to non-European language environments, limiting potential comparison with tonal language environments. Third, infants may have been exposed to other languages in their environment, potentially influencing prosodic behavior. Finally, without video analysis, full functional interpretation of vocal intent could not be performed. Therefore, incorporating behavioral indicators such as eye gaze, gesture, and turn-taking in future studies could clarify the communicative nature of these vocalizations (Papaeliou & Trevarthen, 2006; Wermke et al., 2021).

7. CONCLUSION

In conclusion, this study demonstrates that infants as early as 3 months of age systematically modulate prosodic features to distinguish communicative functions. Both Dutch and Swedish infants produced significantly higher mean pitch and wider pitch span when vocalizing requests compared to comments, indicating that intentional prosodic signaling is present well before the onset of canonical babbling. These findings support the idea that pitch serves as a foundation for early communication and that infants can map sound to communicative functions during the preverbal stage.

While no overall differences emerged between language groups in mean pitch, a significant interaction effect revealed that Dutch infants modulated pitch span more strongly than Swedish infants when distinguishing between requests and comments. This suggests that although prosodic production follows universal developmental trajectories at 3 months, early signs of language-specific prosodic development may already begin to surface in certain acoustic dimensions, particularly pitch variability. These findings highlight the interplay between universal developmental patterns and language-specific input in shaping early vocal behavior, and emphasize the need to study this area of research even further.

0.1 ETHICAL STATEMENT

Ethical approval for the data collection was obtained by the original project (van Niekerk, Junge and Schwarz; 2025). The present study is a secondary analysis of a subset of an existing dataset and follows the methodology outlined in a pre-registered protocol (van Niekerk et al., 2024). All procedures involving human participants were carried out in accordance with the ethical standards of Utrecht University. Informed consent was obtained from the caregivers of all participating infants by the original researchers prior to data collection. The dataset was fully anonymised before being shared with the authors, and no identifying information was accessible during the current analysis.

0.2 ACKNOWLEDGMENTS

Conceptualization: Julia Heemskerk, Patricia Dworak, Yasmina Forsch
Methodology: Elanie van Niekerk, dr. Caroline Junge, and prof. dr. Aoju Chen
Formal analysis: Julia Heemskerk
Investigation: Julia Heemskerk, Patricia Dworak, Yasmina Forsch, Elanie van Niekerk, dr. Caroline Junge, and prof. dr. Aoju Chen
Resources: Julia Heemskerk, Patricia Dworak, Yasmina Forsch
Data Curation: Elanie van Niekerk, dr. Caroline Junge, and prof. dr. Aoju Chen
Writing - Original Draft - Julia Heemskerk
Writing - Review & Editing: Yasmina Forsch, Julia Heemskerk, Patricia Dworak
Visualization: Julia Heemskerk
Supervision: Elanie van Niekerk
Funding acquisition: Elanie van Niekerk, dr. Caroline Junge, and prof. dr. Aoju Chen

The authors would like to thank Elanie van Niekerk, dr. Caroline Junge, and prof. dr. Aoju Chen for granting access to the dataset and for their valuable feedback on the design and interpretation of the present study.

REFERENCES

Astruc, L., Payne, E., Post, B., Vanrell, M. del M., & Prieto, P. (2013). Tonal targets in early child English, Spanish, and Catalan. *Language and Speech, 56*(2), 229–253. https://doi.org/10.1177/0023830912460494

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boersma, P. & Weenink, D. (2025). *Praat: doing phonetics by computer* (Version 6.4.35) [Computer program]. Retrieved June 15, 2025, from http://www.praat.org/

Bruce, G. (2007). Components of a prosodic typology of Swedish intonation. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes: Volume 1: Typological studies in word and sentence processing* (Vol. 1) (pp. 113–146). Mouton de Gruyter. https://doi.org/10.1515/9783110207569

Buder, E. H., Warlaumont, A. S., & Oller, D. K. (2013). An acoustic phonetic catalog of prespeech vocalizations from a developmental perspective. In B. Peter & A. A. N. MacLeod (Eds.), *Comprehensive perspectives on child speech development and disorders: Pathways from linguistic theory to clinical practice* (pp. 103–134). Nova Science Publishers.

Chen, A. (2011). Tuning information packaging: Intonational realization of topic and focus in child Dutch. *Journal of Child Language, 38*(5), 1055–1083. https://doi.org/10.1017/S0305000910000541

Esteve-Gibert, N., & Prieto, P. (2013). Prosody signals the emergence of intentional communication in the first year of life: Evidence from Catalan-babbling infants. *Journal of Child Language, 40*(5), 919–944. https://doi.org/10.1017/S0305000912000359

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60*(6), 1497–1510. https://doi.org/10.2307/1130938

Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child development, 64*(3), 657–674. https://doi.org/10.2307/1131209

Gervain, J., Christophe, A., & Mazuka, R. (2021). Prosodic bootstrapping. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 563–573). Oxford Handbooks. https://doi.org/10.1093/oxfordhb/9780198832232.013.36

Gosselke Berthelsen, S. (2021). The perceptual importance of falling pitch for speakers from different dialects of Swedish. In *Fonetik 2021, 56*, 7–13. Center for Language and Literature, Lund University. Retrieved from https://journals.lub.lu.se/LWPL/article/view/23266/20725

Haan, J. (2002). *Speaking of questions: An exploration of Dutch question intonation*. LOT Publications. Retrieved from https://www.lotpublications.nl/Documents/052_fulltext.pdf

Han, M., & Kager, R. (2020). Pitch properties of infant-directed speech specific to word learning contexts: A cross-linguistic investigation of Mandarin Chinese and Dutch. *Journal of Child Language, 47*(1), 85–111. https://doi.org/10.1017/S0305000919000813

Han, M., De Jong, N. H., & Kager, R. (2022). Prosodic input and children's word learning in infant- and adult-directed speech. *Infant Behavior and Development, 68*, article 101728. https://doi.org/10.1016/j.infbeh.2022.101728

Jhang, Y., & Oller, D. K. (2017). Emergence of functional flexibility in infant vocalizations of the first 3 months. *Frontiers in Psychology, 8*, article 300. https://doi.org/10.3389/fpsyg.2017.00300

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*(23), 1994–1997. https://doi.org/10.1016/j.cub.2009.09.064

Martínez-Álvarez, A., Benavides-Varela, S., Lapillonne, A., & Gervain, J. (2023). Newborns discriminate utterance-level prosodic contours. *Developmental Science, 26*(2), article e13304. https://doi.org/10.1111/desc.13304

Oller, D.K. (2000). *The Emergence of the Speech Capacity* (1st ed.). Psychology Press. https://doi.org/10.4324/9781410602565

Papaeliou, C. F., & Trevarthen, C. (2006). Prelinguistic pitch patterns expressing 'communication' and 'apprehension'. *Journal of Child Language, 33*(1), 163–178. https://doi.org/10.1017/S0305000905007300

Posit team. (2025). *RStudio: Integrated development environment for R* (Version 4.5.3.) [Computer software]. Retrieved April 1, 2026, from https://www.posit.co/

Riad, T. (2014). *The phonology of Swedish*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199543571.001.0001

Romøren, A. S. H. (2016). *Hunting highs and lows: The acquisition of prosodic focus marking in Swedish and Dutch* [Doctoral dissertation, Utrecht University]. LOT Dissertation Series. Retrieved from http://hdl.handle.net/1874/338040

The Language Archive. (2024). *ELAN* (Version 6.9) [Computer software]. Max Planck Institute for Psycholinguistics. Retrieved April 1, 2026, from https://archive.mpi.nl/tla/elan

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science, 11*(3), 188–195. https://doi.org/10.1111/1467-9280.00240

van Niekerk, E., Junge, C., & Chen, A. (2024). *Role of innate mechanisms in the acquisition of prosodic form-meaning mappings. #169759* AsPredicted [Preregistration]. Retrieved from https://aspredicted.org/pe4hc.pdf

van Niekerk, E., Junge, C., & Chen, A. (2025). The Influence of Prenatal Language Exposure on the Use of Pitch in Newborns' Vocalizations: A Systematic Review. *First Language, 46*(2), 305–334. https://doi.org/10.1177/01427237251372203

van Niekerk, E., Junge, C., Schwarz, I. C., Gustavsson, L., Marklund, E., & Chen, A. (2025). *Before Words: How innate knowledge shapes preverbal infants' use of prosody to express communicative functions.* Dag van de Fonetiek 2025, Utrecht, The Netherlands. https://nvfw.org/sites/nvfw.org/files/DvdF25_VanNiekerk_etal.pdf

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wermke, K., Robb, M. P., & Schluter, P. J. (2021). Melody complexity of infants' cry and non-cry vocalizations increases across the first six months. *Scientific Reports, 11*(1), article 4137. https://doi.org/10.1038/s41598-021-83564-8

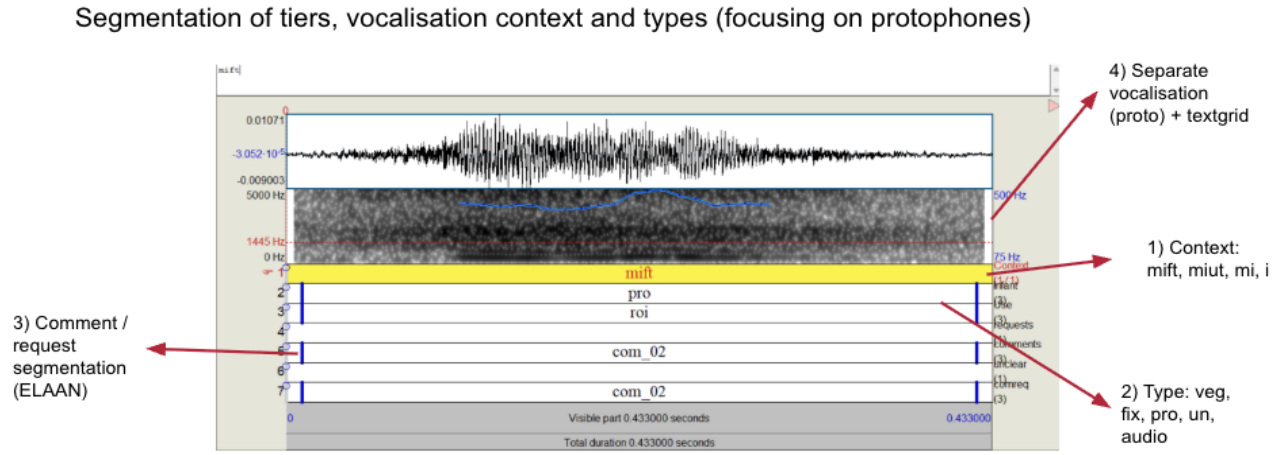
Xu, Y. (2013). ProsodyPro — A tool for large-scale systematic prosody analysis. In B. Bigi & D. Hirst (Eds.), *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (pp. 7–10).

APPENDIX

A. Models mean pitch and pitch span

For mean pitch:
Model A1: f0_mean ~ 1 + (1 | infant)
Model A2: f0_mean ~ function_type + (1 | infant)
Model A3: f0_mean ~ function_type + language + (1 | infant)
Model A4: f0_mean ~ function + language + function*language + (1|infant)

For pitch span:
Model B1: pitch_span ~ 1 + (1 | infant)
Model B2: pitch_span ~ function_type + (1 | infant)
Model B3: pitch_span ~ function_type + language + (1 | infant)
Model B4: pitch_span ~ function + language + function*language + (1|infant)

The final model (step 4) for mean pitch (f0_mean) was model A4: f0_mean ~ function_type * language + (1 | infant). For pitch span it was model B4: pitch_span ~ function_type * language + (1 | infant). The structures of the models allowed us to systematically assess whether communicative function, language, or their interaction significantly influences the pitch characteristics of 3-month-old infants' vocalizations.

B. Example segmentation