Identifying common phrase-level pitch contours in natural infant-directed speech

Manash K. Sahoo[1], Jeremy I. Borjon[1], Chen Yu[1,2], Linda B. Smith[1,3]

[1] Department of Psychological and Brain Sciences, Indiana University, Bloomington, [2] Department of Psychology, University of Texas, Austin, [3] School of Psychology, University of East Anglia

Introduction

Compared to adult-directed speech, infant-directed speech (IDS) is characterized by a higher fundamental frequency, increased pitch variability, exaggerated and repetitive intonation contours, a slower rate of speech, and a distinct spectral timbre[1]. These exaggerated dynamics are thought to increase perceptual saliency of speech and thus facilitate language acquisition[2]. Additionally, the exaggerated dynamics of IDS may modulate the infant's attentional state[1,3], and highlight information at the level of words and sentences[4].

Previous research[5] has identified 4 basic intonation contours (rise, fall, valley, and hill) at the level of individual words during IDS. In natural environments however, children experience language at the level of phrases. In this study, we aim to build upon these results by examining intonation contours on a phrase-level by leveraging a large longitudinal dataset consisting of 17,670 utterances of IDS and using agglomerative hierarchical clustering to organize contours by shape.

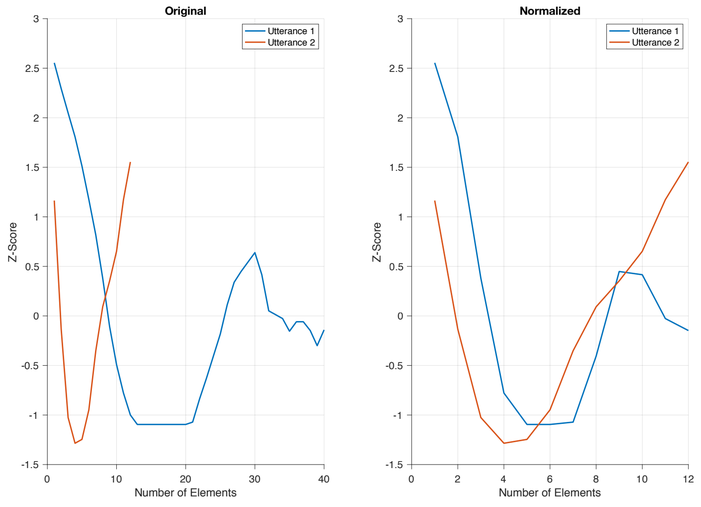
Method

Dataset

Data were taken from 42 unique caregivers engaged in free-play with infants aged 9-24 months old, over 164 experimental sessions. Audio taken from these sessions was recorded at a sampling rate of 16000Hz.
Utterances were manually coded and defined as pauses in speech greater than 400ms.

Data Processing

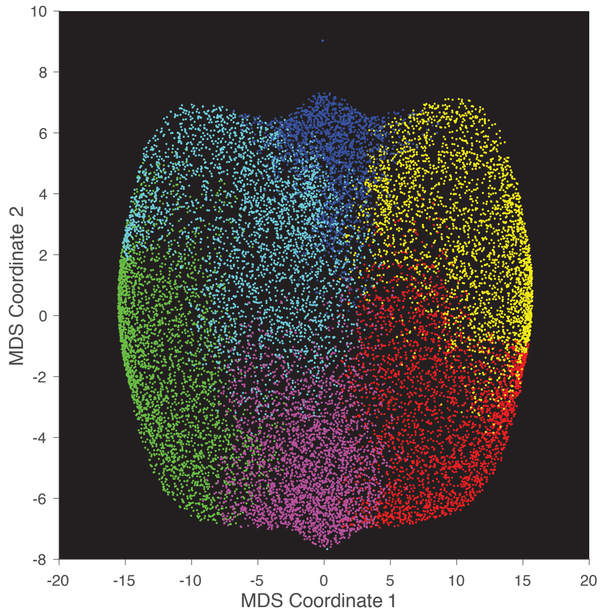
In order to obtain the fundamental frequency from coded utterances, the MATLAB function "pitch" was used, using the cepstrum method.
Pitch contours were then normalized for length, using a ratio of their lengths as a sampling interval (see below). To account for variability in fundamental frequency and noise, pitch contours were z-scored and smoothed using cubic-spline interpolation.



Method

Clustering

Following smoothing, z-scoring, and length-normalization, we used 'complete' agglomerative hierarchical clustering to cluster each of the pitch contours. We then used gap analysis to identify the optimal *k* number of clusters.

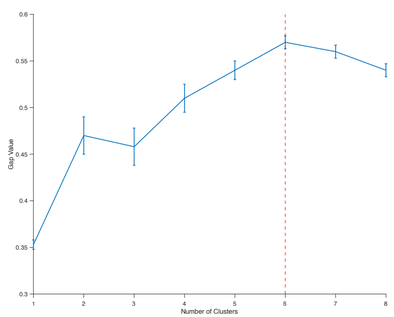


Validation

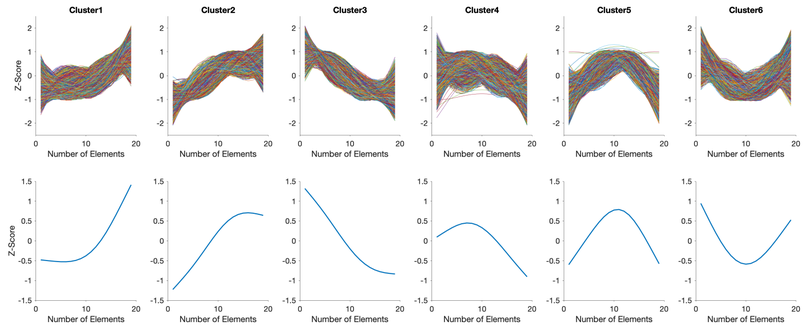
To validate that our clusters were mapped accurately, we trained a k-nearest neighbors clustering algorithm on 80% of pitch contours, leaving 20% as testing data. With this method, we achieved 97.548% accuracy.

Results

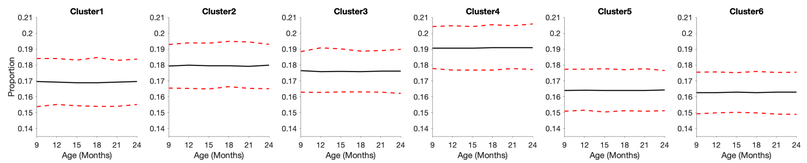
To verify the optimal number of clusters, gap-analysis was used. The maximal gap value appeared at K = 6 clusters.



Results



We found 6 different clusters of intonation curves. In addition to hill and valley type contours (clusters 5 and 6), we found contours that would initially rise or fall and gradually flatten or begin flat and gradually rise or fall.



Additionally, we found that the proportion of these clusters of caregiver speech did not vary significantly across ages of development.

Conclusion

Using agglomerative hierarchical clustering, we clustered the pitch contours of infant-directed caregiver speech into six different categories, where contours would initially rise or fall and gradually flatten, begin flat and gradually rise or fall, or follow a 'hill' and 'valley' type shape. Mapping an 80/20 training-test split of these clusters to a K-Nearest Neighbors algorithm resulted in 97.548% accuracy, suggesting that these clusters are highly consistent. Furthermore, we found no significant variation in the proportion of these clusters throughout 9-24 months of development. Further research will examine intonation contours on a word-level, and how these clusters of caregiver speech modulate infant behavior such as sustained attention and object exploration during naturalistic play.

References

1. Fernald, A. (1992). *Meaningful melodies in mothers' speech to infants.* In H. Papoušek, U. Jürgens, & M. Papoušek (Eds.), *Studies in emotion and social interaction. Nonverbal vocal communication: Comparative and developmental approaches* (p. 262–282). Cambridge University Press; Editions de la Maison des Sciences de l'Homme.
2. Thiessen, E., Hill, E., & Saffran, J. (2005). Infant-Directed Speech Facilitates Word Segmentation. *Infancy, 7(1), 53-71.*
3. Kuhl, P.K. (2007), Is speech learning 'gated' by the social brain?. Developmental Science, 10: 110-120.
4. Fernald A. Intonation and communicative intent in mothers' speech to infants: is the melody the message? Child Dev. 1989 Dec;60(6):1497-510.
5. Nencheva ML, Piazza EA, Lew-Williams C. The moment-to-moment pitch dynamics of child-directed speech shape toddlers' attention and learning. Dev Sci. 2021 Jan;24(1):e12997.