Interspeech 2022
18-22 September 2022, Incheon, Korea

Fundamental Frequency Variability over Time in Telephone Interactions

Leah Bradshaw[1], *Eleanor Chodroff*[2], *Lena Jäger*[1], *Volker Dellwo*[1]

[1]Department of Computational Linguistics, University of Zurich, Switzerland
[2]Department of Language and Linguistic Science, University of York, UK
leah.bradshaw@uzh.ch

Abstract

Speech signals contain substantial fundamental frequency (f_0) variability. Even within a single utterance, speakers modify f_0 to create different intonational patterns. Previous studies have identified markers of increased f_0 variability, such as the introduction of a new topic or greetings, but these are limited in the scope of their analyses. In the present study, we investigate f_0 variability over the course of a telephone conversation, with a focus on the initial and medial utterances within the exchange. We examined f_0 standard deviation of each utterance in over 2000 telephone conversations from 509 American English speakers from the Switchboard corpus. Findings showed that on average, speakers exhibit more f_0 variability in the opening compared to mid-conversation utterances. Further, findings suggest that the inclusion of a greeting word in an initial turn, e.g., "hello" or "hi", corresponds to an increase in f_0 standard deviation. These results suggest that speakers employed more variable f_0 in the initial few turns of a telephone conversation. The interpretation of this finding is multifaceted and may be linked to several communicative goals, including the placement of identity markers in conversation or the attraction of attention, or the role of openings as boundary markers.

Index Terms: fundamental frequency, telephone conversations, discourse structuring

1. Introduction

Fundamental frequency (f_0) varies substantially in a single utterance as a result of various linguistic, paralinguistic, and non-linguistic parameters. In this study, we investigated whether the amount of f_0 variability in an utterance also relates to its position in time in telephone interactions. It is known that f_0 variability is greater at intonational phrase boundaries within a given utterance, and that f_0 range is expanded to mark speech acts within discourse above the utterance level. However, there are substantial limitations to these findings with regards to telephone openings, and particularly greetings. This study seeks to validate whether boundary marking and the conversational goals in the opening sequences of telephone conversations lead to marked f_0 variability. Our focus is on telephone openings and greetings within telephone openings, as literature frequently shows interest in these, citing greater f_0 variability [e.g., 1–3], but acoustic or quantitative validation has been somewhat limited (see section 1.2). We present here a large-scale analysis of f_0 variability using over 500 American English speakers from the Switchboard corpus [4].

1.1. Fundamental frequency variability

In speech, fundamental frequency (f_0) is an acoustic measure which reflects the rate of vocal fold vibration and correlates with the perceived pitch of the speaker [5]. Variability in f_0 within an utterance is shown to correspond to a range of parameters. From a linguistic perspective, marking of illocutionary force is frequently established by means of f_0 modifications [e.g., 6–8]. For English, intonational phrase boundaries are expressed with phrase accents and boundary tones that may increase f_0 variability. Moreover, the final, nuclear pitch accent in an intonational phrase may also convey discourse-level information and pragmatic meaning [9–11]. More recent studies suggest that prenuclear pitch accents may also be relevant for conveying discourse information such as givenness [12–13]. In any case, the use of pitch accents frequently corresponds to increased f_0 variability, thereby conveying structure and meaning simultaneously.

Moving slightly above the utterance, f_0 modifications also reflect turn-structuring in conversations [e.g., 14–15]. Moreover, a few early studies examined prosodic boundaries over multiple utterances that could reflect higher-level discourse structure. In particular, f_0 ranges are greater in utterances which signal the beginning of a new topic [16–21], and lower f_0 values are observed at topic boundaries [22]. Further, discourse segment boundaries are shown to be marked by higher variation in f_0 [23–24]. One shared feature of these within- and between-utterance f_0 modifications is their relationship to boundaries, where greater modifications are typically associated with the occurrence of a boundary.

Beyond structure building, f_0 variability is also shown to relate to some paralinguistic parameters, including emotion and attitude [e.g., 25–27], and non-linguistically with speaker age and gender differences [28].

1.2. Telephone openings and greetings

Telephone dialogues offer a relatively structured form of discourse; however, the lack of visual information in this medium requires callers to rely more heavily on verbal cues for efficient transfer of information. It is well-acknowledged that the organisational structure of telephone conversations does not differ much from a typical face-to-face conversation; namely, they involve an opening and closing with some talk in between [2, 29–31]. Within this sequence, openings in particular offer an interesting point of analysis given their consistent role as a topic boundary and frequent inclusion of greetings.

Early studies argue that pitch variability occurs in telephone openings, perhaps as a result of speakers attempting to identify one another in initial turns [1–2]. However, these studies were completed impressionistically with no acoustic or quantitative validation of these claims.

Indeed, further research has investigated the acoustic properties of various speech acts, with some minor exploration into greetings. It is acknowledged that speech act category, e.g., question, statement or greeting, strongly influences the prosodic structure of an utterance, particularly in relation to f_0 mean and range [32–33]. Limited findings show that greetings in

10.21437/Interspeech.2022-10669

particular exhibit uniformly larger f_0 ranges compared to many other speech acts [33]. However, this finding was established using data from only one speaker in scripted speech, giving no indication as to whether it might generalise across speakers or to spontaneous naturalistic speech.

A preliminary investigation into the specific acoustic properties of different greetings, namely types of "hello", as they occur in telephone openings has been offered by [3]. Using 64 calls, f_0 contours were analysed to explore the differences between stand-alone "hello" tokens, and those which contained "hello" followed by more talk. "Hello" in the former signals a full turn, whilst in the latter signals that the current speaker will continue with their turn. The author noted different f_0 contours in the word "hello" depending on this role, however, these did not appear to correlate with signalling incomplete turns where more talk would follow the "hello". Over half of all turns containing a stand-alone "hello" and the majority of multi-unit turns containing an initial "hello" had an overall rising f_0 contour, characterised by a rise mid-to-high in the speakers' range in the second syllable of the lexical item. More than half the duration of these tokens was taken up by the last vowel which was also the locus of the f_0 dynamics. Although this study offers a thorough analysis of f_0 contours in "hello" in telephone calls, this analysis is limited to the differences between individual greeting types and does not consider their relationship to subsequent utterances in the call produced by the same speaker. Further, this study focuses solely on familiar speakers, so it is unclear if this f_0 variability is also present in calls between unfamiliar speakers.

The findings from the above studies all suggest that telephone openings may exhibit greater f_0 variability due to their purpose for greetings, with some limited acoustic evidence supporting this. Examination into the communicative goals of greetings can equally motivate the existence of this greater f_0 variability in these utterances.

For instance, the lexical item "hello", and cultural equivalents, have been used to explore personality ratings [34–36]. Interestingly, [35–36] show a great degree of consistency in listener ratings of personality traits even from such short samples. We could reasonably suggest that some level of additional speaker-specific information is being conveyed here that allows for the consistency in these findings. Perhaps speakers are intentionally expressing personality in these opening utterances or aiming to express features such as warmth or likeability. Further, these studies also cite f_0 measures of mean, range (f_0 *max* – f_0 *min*) and glide (f_0 *end* – f_0 *start*) as a cue to personality [35–36]. This could suggest that f_0 variability in greetings may be a function for transferring speaker-specific information, such as personality.

1.3. Variability as a function of discourse

Moreover, some previous findings suggest that sweeping f_0 contours lead to improvements in human and machine speaker identification performance [37]. Therefore, increased f_0 variability may be present in telephone openings more generally, including utterances not containing greetings, to assist the callee with recognising them. It is possible that when answering calls, additional identity-specific vocal cues accompany the talk to assist with the necessary confirmation of speaker identity. Potentially, these vocal cues may correspond acoustically to these f_0 modifications which assist speaker identification.

Additionally, speakers may be using attention grabbing devices to ensure they have the full attention of their interlocutor prior to engaging in the main purpose of the call. This attention grabbing may be accompanied by similar acoustic properties to what is seen in infant-direct speech interactions [e.g., 38–40], which would lead to greater f_0 variability. Regardless of speaker intentions, there is sufficient evidence to suggest that f_0 variability is a key component in opening utterances in calls.

1.4. Research hypotheses

This study investigates f_0 variability from opening to mid-conversation utterances in telephone interactions. We expect increased f_0 variability in the opening turns of a call relative to later points. Findings have suggested that greetings in particular contain higher f_0 variability, however, more generally, telephone openings may be marked in terms of f_0 variability as a result of their role as a discourse topic marker, or the speakers' intentions for identification and attention grabbing. Further, previous studies are limited in their analyses of f_0 modifications in greetings, with verification of these claims somewhat lacking in this body of research. This study employs large-scale acoustic analysis and modern statistical methods to quantify the extent of f_0 variability across speakers.

2. Data and methods

2.1. Materials

The analysis used recorded telephone conversations between American English speakers taken from the Switchboard-1 Telephone Speech Corpus [4]. The recordings consisted of telephone interactions between unfamiliar speakers, which were monitored by a computer-driven robot operator. This robot operator was responsible for giving the caller a recorded prompt; selecting and dialling the callee; introducing the topic for discussion; and recording the speech from two subjects into separate channels. Selection of callers within the study was constrained so that no two speakers would converse more than once, such that all calls were between two unfamiliar speakers. Speakers had the option to "warm up", meaning some recordings contain a greeting interaction.

For the purpose of this analysis, stereo channel calls were divided into mono channels, such that they contained a single speaker. In the initial extraction, 4700 call-sides were used, containing 518 speakers in total. However, channels with fewer than 20 suitable utterances were then excluded from the analysis, assumed not to successfully represent the full scope of a typical telephone conversation, resulting in 4368 call-sides. The final analysis used speech from a total of 509 speakers (270 male, 239 female).

2.2. Measurement

Utterance-level alignments accompanied the recordings which were used to automatically extract f_0 standard deviations in ERB from each utterance in Praat [41]. ERB corresponds to equivalent rectangular bandwidth and serves as a psychoacoustic transformation of hertz [42]. Measures were taken using gender-specific pitch ranges (50–200Hz for males; 75–400Hz for females). The duration (ms) of each utterance was also calculated to account for variability in utterance length. Utterances which had been marked as containing "noise" (i.e., background noise, unintelligible speech or phone noise) were removed from the analysis.

Our main goal involved determining whether utterances in the initial- and medial-time course of a telephone conversation contain different amounts of f_0 variability. Therefore, only utterances to the midpoint of the suitable utterances for each speaker in each call were used, corresponding to a substantial portion of the first half of the interaction. Given that few calls contained more than 60 suitable utterances, the midpoint was capped at 30 utterances, such that speakers who produced more than this in the first half of the call were only represented by their first 30.

2.3. Statistics

In order to model the findings of this experiment, we used Generalised Additive Mixed-Models (GAMMs) [43]. This allowed for a model representing the non-linear and time-series relationship in the results. All models were fitted using the mgcv::bam function in R [43] and the itsadug::compareML function was used for model comparisons.

We fitted a GAMM to explore the effect of the position of the utterance within the call (*utterance number*) on the amount of f_0 variability (*standard deviation*). The model estimated f_0 standard deviation from parametric terms of utterance duration (dur), gender and the interaction between utterance number (utt) and duration. Smooth terms were fitted for utterance number, as well as a random effect of speaker and utterance number by speaker.

Following the recommended procedure in [44], the full model was compared with a nested model which excluded all terms for the predictor being tested (utterance number).

3. Results

The model revealed significant effects of the parametric term *duration*, as well as a significant interaction between *utterance number* and *utterance duration* (each $p < 0.0001$). The smooth term parameter for *utterance number* was also statistically significant, showing the relationship between f_0 standard deviation and *utterance number* is highly non-linear. Table 1 shows the summarised GAMM output for the full model.

Table 1: *Summary of full GAMM model*

Parametric Coefficients	**Estimate**	***t*-value**		***p*-value**
(Intercept)	0.56	268.95		*<0.001*
duration	0.02	61.32		*<0.001*
gender-Female	0.14	131.00		*<0.001*
Smooth Terms	**edf[1]**	**Res.df**	***F***	***p*-value**
s(utterance number (#))	8.38	8.88	45.02	*<0.001*
ti(utterance #, duration)[2]	10.62	11.70	50.31	*<0.001*
s(speaker, utterance #)	<0.0001	1.00	0.00	*<0.01*
s(speaker)	2.73	1.00	26.07	0.198

According to the model comparison of the full and nested model, the inclusion of *utterance number* as a predictor within the model significantly improves the model fit ($p < 0.0001$), which corresponds to a significant difference in the shape of the slopes of the two models.

Visualisation of the full GAMM model (Figure 2) shows that the initial few utterances in the call have substantially more f_0 variability, followed by a sharp decrease until the 5th utterance. Notably, we also see a relatively tight confidence interval in this area of the curve, suggesting little between-speaker variability for this tendency to exhibit higher f_0 variability in the first few utterances. f_0 variability then remains relatively stable for the rest of the utterances, but we see greater between-speaker variability.

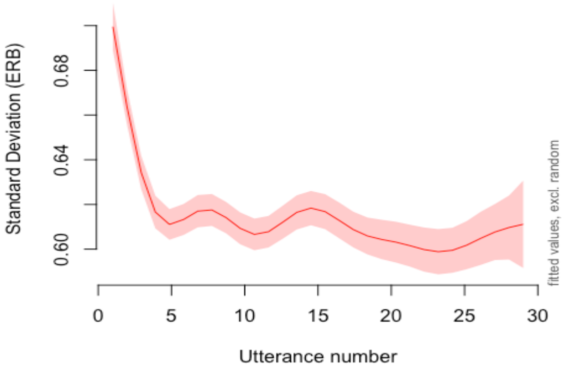

Figure 1: *Nonlinear smooth fitted for f_0 standard deviation in each utterance across a call. Shaded bands represent the pointwise 95%-confidence interval*

To consider the effect of a greeting term in the utterance on the f_0 standard deviation, a linear mixed effects model was fitted using the lme4 function in R [45]. A binary distinction of '*contains greeting'* or '*doesn't contain greeting*' was marked for each utterance with the inclusion of the following words used to signify '*contains greeting*'; 'hello', 'hi', 'hey', 'good morning' and 'good afternoon'. Only the first 5 utterances were considered for each call, as these are most likely to contain these greeting terms. Further, the f_0 variability of the first 5 utterances follows a linear pattern, allowing for use of a linear model. The binary distinction of *contains greeting* was included in the model as a predictor, along with gender, an interaction between *utterance number* and *duration* and a by-speaker random effect. The model effects are presented in Table 2; all predictors were significant, but of particular interest for the purpose of this analysis are *contains greeting* and *utterance number*.

Table 2: *Summary of full LME model*

Fixed Effects	**Estimate**	***t*-value**	***p*-value**
(Intercept)	0.072	71.77	*<0.001*
utterance#	-0.0039	-14.26	*<0.001*
duration	0.00023	2.40	*0.0163*
genderMale	-0.028	28.51	*<0.001*
helloYes	0.0095	3.83	*<0.001*
utterance#: duration	0.00032	6.09	*<0.01*

[1] EDF is the effective degrees of freedom used by a smooth given the number of basis functions and the smoothing parameter [44]. Residual degrees of freedom (Res.df) is the number of data minus model degrees of freedom.

[2] This denotes the interaction term between utterance number and duration.

We found that the predictor *contains greeting* corresponded to a significant increase in f_0 standard deviation ($\beta = 0.0095, p < 0.001$). All other predictors were also significant, including utterance number which showed that an increase in *utterance number* corresponded to a statistically significant decrease in f_0 standard deviation ($\beta = -0.0039, p < 0.001$), corroborating the GAMM model output.

4. Discussion

Overall, we see a strong and consistent general trend in these findings for initial utterances to contain substantially more f_0 variability than mid-conversation utterances. Further, our findings show that f_0 variability becomes more stable in mid-conversation utterances; however, there is also more between-speaker variability. We acknowledge here some limitations of this spontaneous speech dataset, such as lack of consistency in the total number of utterances, duration of utterances and content of utterances, which make it difficult to interpret mid-conversation findings. In the following discussion, we will therefore focus on the primary trend for greater f_0 variability in opening utterances and explore some potential interpretations of this finding.

Firstly, our findings support previous evidence that shows that topic boundaries are marked by greater f_0 variability [16–21]. Indeed, telephone conversation openings act as a topic boundary and we see a peak in f_0 variability at this point where the topic boundaries align. In future research, it would be interesting to consider if the opening as a conversational boundary, and therefore the *first* new topic, is somehow differentiated in f_0 variability relative to subsequent new topics within a conversation. Given the scale of our dataset and the fact that new topic introductions occur at different points within the conversations, this particular investigation was beyond the scope of this study.

Further, the shape of the slope, which indicates a rapid decrease in f_0 variability around the 5th utterances, could correlate with theories of prosodic entrainment [cf. 47–48]. Speakers adjust their prosodic features in relation to their respective interlocutor, becoming closer to one another as a conversation progresses. It is possible that speakers converge towards one another and the variability which we see in these initial utterances is speakers manipulating their f_0 to find this convergence level. Whether the two speakers within an individual call indeed have more similar absolute f_0 levels later in the call, however, remains to be seen.

Interestingly, our findings show that openings containing a greeting show a significant increase in f_0 variability compared to openings that did not. This validates previous findings indicating greetings are marked by increased f_0 variability, and further reveals that greetings result in an even greater degree of f_0 variability, compared to utterances which act solely as topic/discourse boundaries. Some potential factors governing this increased f_0 variability in utterances containing greetings are discussed below.

For instance, given that the speakers are unfamiliar with one another in the Switchboard corpus, f_0 variability may arise from some context-specific communicative goals. Speakers may aim to express positive personality attributes and create a positive first impression in these initial turns. Equally, since the participants task is to discuss a specific topic, speakers may want to ensure they have the full attention of their interlocutor prior to this discussion. Previous studies have linked positive valence ratings [35, 46] and attention grabbing [e.g., 38-40] with increased f_0 variability, including in greetings specifically.

Additionally, speakers may enhance cues to their identity through f_0 modifications. Recent findings have suggested that speakers are able to alter their vocal properties to make themselves more recognisable [49]. We can directly link this to the purpose of the opening turns for callers to identify one another. Indeed, findings have shown that f_0 variability in vowels can offer recognition advantages for both human listeners and machines [40]. In a voice discrimination task, results showed performance improvements in vowels which contained sweeping f_0 contours compared with those that had steady state f_0 throughout, suggesting f_0 manipulations may be used by speakers to make themselves more easily recognisable.

The lack of verbal identity cues in telephone conversations mean quick identification is necessary solely through vocal cues. In the modern day, this identification is greatly assisted by Caller ID, however, in times prior to this, it is plausible that speakers may offer this "identity-marked talk" subconsciously in the beginnings of calls. Therefore, although our speakers are unfamiliar with one another, these recordings were made when Caller ID was still relatively new, and this talk could be identity-marked for easier recognition, as suggested by the aforementioned studies.

Naturally, the above represents many hypothetical explanations for the increased f_0 variability in utterances containing greetings. It is likely that a combination of these intentions contributes to the increase in f_0 variability that we observed in greetings. Further, it remains unclear which of the factors discussed contribute to the increased f_0 variability in telephone openings more generally, or greetings specifically.

Additional research is necessary to examine to what extent, if any, these factors have on the amount of variability in these opening utterances. For instance, it would be beneficial to examine calls between familiar speakers to explore if this f_0 variability is beneficial for speaker identification to assess if identity-marking is playing a role here. Equally, interesting extensions to this research could explore the influence of gender-matched vs. mismatched conversations, or the effect of language/culture. Finally, valuable insights may be sought with further investigation into the role of f_0 on a more global level in conversation structuring.

5. Conclusion

The present study investigated how overall f_0 variability differs over time in telephone interactions, focusing on the initial to medial time course within the conversation. Findings showed that opening utterances contained considerably greater f_0 variability compared to mid-conversation utterances. The interpretation of this finding is discussed at length and attributed to many potential factors within- and above the linguistic level, however, we concede that any single or combination of these factors could be at play here. Overall, this study suggests that f_0 may play a bigger role as a function of discourse than already considered, and an exploration into this on a higher level is worthy of greater attention.

6. Acknowledgements

This research was supported by the Swiss National Science Foundation in the project # 185399 "The dynamics of indexical information in speech and its role in speech communication and speaker recognition".

7. References

[1] H. Sacks and E. A. Schegloff, "Two preferences in the organization ofreference to persons in conversation and their interaction," *Everyday Language: Studies in Ethnomethodology. New York*, 1979.

[2] E. A. Schegloff, "Identification and Recognition in Telephone Conversation Openings," *Everyday Language: Studies in Ethnomethodology, New York, Irvington*, pp. 23–78, 1979a.

[3] M. Kaimaki, "Transition relevance and the phonetic design of English call openings," *Journal of Pragmatics*, vol. 43, no. 8, pp. 2130–2147, 2011.

[4] Godfrey, John J., and Edward Holliman. Switchboard-1 Release 2 LDC97S62. Web Download. Philadelphia: Linguistic Data Consortium, 1993.

[5] S. Z. Li and A. Jain, Eds., "Fundamental Frequency, Pitch, F0," in *Encyclopedia of Biometrics*, Boston, MA: Springer US, 2009, pp. 592–592.

[6] J. Coates, "The semantics of the modal auxiliaries: a corpus based analysis with special reference to contemporary spoken English.," Doctoral Dissertation, University of Lancaster, 1980.

[7] M. A. K. Halliday, "Functional Diversity in Language as Seen from a Consideration of Modality and Mood in English," *Foundations of Language*, vol. 6, no. 3, pp. 322–361, 1970.

[8] J. Holmes, "Modifying illocutionary force," *Journal of Pragmatics*, vol. 8, no. 3, pp. 345–365, 1984.

[9] M. A. K. Halliday, *Intonation and Grammar in British English*. De Gruyter Mouton, 1967.

[10] D. R. Ladd, *The structure of intonational meaning: Evidence from English*. Indiana University Press, 1980.

[11] C. Gussenhoven, *On the Grammar and Semantics of Sentence Accents*. De Gruyter Mouton, 1984.

[12] E. R. Chodroff and J. Cole, "Information structure, affect, and prenuclear prominence in American English," in *INTERSPEECH 2018*, Sep. 2018, pp. 1848–1852.

[13] E. R. Chodroff and J. Cole, "The phonological and phonetic encoding of information status in American English nuclear accents," in *Proceedings of the 19th ICPhS,* 2019.

[14] S. Baumann, "Information structure and prosody: Linguistic categories for spoken language annotation," in *Methods in Empirical Prosody Research*, De Gruyter, 2006, pp. 153–180.

[15] P. French and J. Local, "Prosodic Features and the Management of Interruptions 1," in *Intonation in discourse*, Routledge, 1986, pp. 157–180.

[16] I. Lehiste, "The phonetic structure of paragraphs," in *Structure and process in speech perception*, Springer, 1975, pp. 195–206.

[17] E. A. Schegloff, "The relevance of repair to syntax-for-conversation," in *Discourse and syntax*, Brill, 1979, pp. 261–286.

[18] D. Brazil, *Discourse intonation and language teaching.* ERIC, 1980.

[19] B. Butterworth, "Hesitation and semantic planning in speech," *Journal of Psycholinguistic Research*, vol. 4, no. 1, pp. 75–87, 1975.

[20] G. Brown, "Prosodic structure and the given/new distinction," in *Prosody: Models and Measurements*, Springer, 1983, pp. 67–77.

[21] G. M. Ayers, "Discourse functions of pitch range in spontaneous and read speech," 1994.

[22] E. Shriberg, A. Stolcke, D. Hakkani-Tür, and G. Tür, "Prosody-based automatic segmentation of speech into sentences and topics," *Speech Comm.*, vol. 32, no. 1, pp. 127–154, 2000

[23] J. Hirschberg and J. Pierrehumbert, "The intonational structuring of discourse," in *24th Annual Meeting of the ACL*, 1986, pp. 136–144.

[24] B. Grosz and J. Hirschberg, "Some intonational characteristics of discourse structure," 1992.

[25] G. L. Huttar, "Relations between prosodic variables and emotions in normal American English utterances," *JSLHR*, vol. 11, no. 3, pp. 481–487, 1968.

[26] I. Abe, "How vocal pitch works," The Melody of Language, University Park Press, Baltimore, pp. 1–24, 1980.

[27] I. R. Murray and J. L. Arnott, "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion," *JASA*, vol. 93, no. 2, pp. 1097–1108, 1993.

[28] D. Bolinger and D. L. M. Bolinger, *Intonation and its uses: Melody in grammar and discourse*. Stanford university press, 1989.

[29] E. A. Schegloff, "Sequencing In Conversational Openings," in *Selected Studies and Applications*, J. A. Fishman, Ed. De Gruyter Mouton, 1972, pp. 91–125.

[30] E. A. Schegloff and H. Sacks, "Opening up closings," 1973.

[31] G. Jefferson, "A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences," 1973.

[32] E. Shriberg *et al.*, "Can Prosody Aid the Automatic Classification of Dialog Acts in Conversational Speech?," *Lang Speech*, vol. 41, no .3–4, pp .443–492, Jul.1998.

[33] A. K. Syrdal and Y.-J. Kim, "Dialog speech acts and prosody: Considerations for TTS," *Speech Prosody*, p. 5, 2008.

[34] C. Baus, P. McAleer, K. Marcoux, P. Belin, and A. Costa, "Forming social impressions from voices in native and foreign languages," *Sci Rep*, vol. 9, no. 1, p. 414, Dec. 2019.

[35] P. McAleer, A. Todorov, and P. Belin, "How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices," *PLoS ONE*, vol. 9, no. 3, p. e90779, Mar. 2014.

[36] C. Ferdenzi, S. Patel, I. Mehu-Blantar, M. Khidasheli, D. Sander, and S. Delplanque, "Voice attractiveness: Influence of stimulus duration and type," *Behavior Research Methods*, vol. 45, no. 2, pp. 405–413, Jun. 2013.

[37] V. Dellwo, T. Kathiresan, E. Pellegrino, L. He, S. Schwab, and D. Maurer, "Influences of Fundamental Oscillation on Speaker Identification in Vocalic Utterances by Humans and Computers," in *INTERSPEECH 2018*, Sep. 2018, pp. 3795–3799.

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

[39] 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," *Journal of Child Language*, vol. 16, no. 3, pp. 477–501, 1989.

[40] B. Gauthier and R. Shi, "A connectionist study on the role of pitch in infant-directed speech," *JASA*, vol. 130, no. 6, pp. EL380–EL386, 2011.

[41] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]," *Version 6.2, retrieved from http://www.praat.org/*, 2021.

[42] B. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hearing research*, vol. 47, no. 1–2, pp. 103–138, 1990.

[43] S. N. Wood, Generalized additive models: an introduction with R. CRC press, 2017.

[44] M. Sóskuthy, "Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction," *arXiv:1703.05339 [stat]*, Mar. 2017, Accessed: Nov. 19, 2021. [Online].

[45] D. Bates, M. Maechler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software*, 67(1), 1-48, 2015.

[46] B. L. Brown, W. J. Strong, and A. C. Rencher, "Fifty-four voices from two: the effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech," *JASA,* vol. 55, no. 2, pp. 313–318, Feb. 1974

[47] R. Levitan and J. B. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," 2011.

[48] R. Levitan, A. Gravano, L. Willson, Š. Beňuš, J. Hirschberg, and A. Nenkova, "Acoustic-prosodic entrainment and social behavior," in *NAACL HLT'12*, 2012, pp. 11–19.

[49] V. Dellwo, E. Pellegrino, L. He, and T. Kathiresan, "The dynamics of indexical information in speech: Can recognizability be controlled by the speaker?," *AUC PHILOLOGICA*, vol. 2019, no. 2, pp. 57–75, 2019.