

Supporting Information for

Structure in conversation: Evidence for the vocabulary, semantics and syntax of prosody

Nadav Matalon^{1,2,†}, Eyal Weinreb^{1,†}, Dominik Freche¹, Erez Volk³, Tirza Biron⁴, Elisha Moses^{1,*}, David Biron^{5,*}

¹ Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel.

² Department of Linguistics, Hebrew University of Jerusalem, Jerusalem, Israel.

³ NeuraLight Inc., Tel Aviv, Israel.

⁴ Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel.

⁵ The Data Science Institute and The Department of Statistics, The University of Chicago, Chicago, IL, USA.

* Corresponding authors: elisha.moses@weizmann.ac.il (EM), dbiron@statistics.uchicago.edu (DB)

†These authors contributed equally to this work

This PDF file includes:

SI Text S1 to S5
Figures S1 to S11
Tables S1 to S10
SI References

Other supporting materials for this manuscript include the following:

Supplementary audio S1 to S4

Text S1 - Rationale for the search for a prosodic vocabulary

Prosody functions as a simple structured communicative system, much like a language. Analogous to a language, prosody comprises a set of fundamental units and a system of rules governing their combination (syntax and/or grammar). Our findings identify these characteristics in natural conversations. The notion of a "prosodic vocabulary" is a main thrust of this paper. Since the inventory of pitch contours that we identify serves as a set of elements that combine to form prosodic structures, we draw the analogy to the construction of linguistic utterances.

This analogy is consistent with linguistic definitions: Language involves a system of communication that utilizes signs (or symbols) to convey a meaning. These signs are the elements of the vocabulary, and their full set is the "vocabulary" of the language. The language then employs a set of rules for concatenating the basic elements into larger segments called sentences. These rules comprise a syntax, and a grammar for constructing the variety of structures and sentences.

To identify the prosodic vocabulary, we used clustering. A vocabulary element is a concrete object representing a cluster - the centroid pitch contour. Figs. S3, S6, S8, and S10 show the representative pitch contours of the clusters, and these are the conceptual elements of the prosodic vocabulary. This concretization, while it is still an analogy, demonstrates what we mean by a prosodic vocabulary.

Text S2 - Prosodic vocabulary vs verbal vocabulary

An intriguing question that was not systematically explored in this study is the comparison between the well-established verbal vocabulary and the prosodic vocabulary.

One aspect is the relation between form and function in the vocabulary. We observed that, within the scope of our dataset, similar pitch contours can convey different meanings (serve distinct functions). At the same time, certain structural features, such as large pitch movements or a high register, appear to carry overlapping meanings, often associated with emphasis or strong emotion. Fig. S6 presents two examples where similarities in form coincide with similarities in meaning. A parallel can be drawn to verbal language, where polysemy is common. English words such as 'love', 'play', or 'answer' function as both nouns and verbs, while others, like 'bank', 'spring', or 'date' can convey multiple unrelated meanings. Conversely, distinct words can express highly similar meanings. In both verbal and prosodic domains, correct interpretation ultimately depends on contextual cues. Nevertheless, analyzing prosodic building blocks does suggest that they are more fluid than English words.

A main difference between prosody and text lies in the type of meaning which is conveyed. Unlike the meaning of words, prosodic meaning is typically not referential, i.e., pointing to some entity, object, action, or concept which exists outside the linguistic realm. Prosody functions in conversation as a contextualization device (1, 2), that is, creating a correct frame of interpretation for the words which carry it. Thus, depending solely on prosody, a responsive utterance "what?" could signal that prior talk was not properly perceived and should be repeated (small pitch rise), or that it counters expectations and should be explained (large pitch rise) (3).

In summary, our findings agree with the widely recognized understanding that prosody exhibits a higher degree of ambiguity and redundancy compared to the verbal vocabulary of a language. In the latter, word distinctions are generally more categorical and less variable in interpretation. Fully addressing this comparison is challenging due to the interdependence of prosody with both the textual tier and the broader conversational context. It probably should begin with a systematic characterization and categorization of the meaning for a large subset of pitch contours. As such, a comprehensive resolution of this question lies beyond the scope of the present study.

Text S3 - Previous approaches to define a prosodic vocabulary

The two major theoretical approaches to define a vocabulary of English prosody are constructive and “top-down”. These approaches build the full repertoire of pitch contours from a limited inventory of basic constituents. The constituents combine with each other by occupying specific locations within the IU.

In the “British school of intonation” (4–7), the basic constituents are “tones”, i.e., micro pitch movements. Tones may occupy four possible parts within a well-formed IU: an obligatory nucleus, and optional tail, head, and pre-head. The possible combinations give rise to approximately 100 contours. However, the practical size of the English intonational “vocabulary” is typically estimated as a few dozen commonly used and meaningful “tunes” (5, 6).

The second approach is the “Autosegmental-Metrical theory of Intonation” (8–10). In contrast to the aforementioned “tone” approach, pitch contours are viewed here as a series of individual events, between which intonation is transitional and phonologically unspecified. The basic constituents are target “levels” that coincide only with accented syllables (“pitch accents”) and unit boundaries (“edge tones”). The ToBI annotation system (11) is a widely-accepted formalization of this theory, used as a means for representation of pitch contours. According to this system, American English exhibits an inventory of ten distinct edge tones and pitch accents (some of which exhibit a down-stepped variant) (12). The assumption is that an IU may exhibit up to three pitch accents and two edge tones, and that all sequences of pitch accents and edge tones are well-formed (8). Thus, this theory suggests an order of magnitude of >1,000 distinct contours.

In contrast, our approach to identifying the building blocks of the prosodic vocabulary is bottom-up, without assuming a predetermined set of constituents or an inner structure of the IU. However, this does not imply that the identified patterns lack an internal structure, nor does it preclude describing them using the terms of the British school system of “tones” or the ToBI annotation system. Fig. S8 demonstrates this by showing translations of the prosodic patterns described in Fig. 2b-c into these two other systems.

Text S4 - Understanding our analysis of functions

The multilayered functions that prosody serves in conversation were tackled from several perspectives. One approach differentiates linguistic and emotional (“paralinguistic”) meanings, and focuses solely on the former while disregarding the other (10, 13).

Our analysis follows a different trajectory, drawing upon the theoretical foundations of functional linguistics (14–16). Building on this tradition, research on prosody in conversation adopts a holistic and integrative perspective (17–19). Within this framework, all aspects of meaning that can be attributed to variations in suprasegmental vocal features - such as pitch, loudness, timing, and voice quality - are considered prosodic, i.e., integral to the linguistic system, provided they are demonstrably relevant to the interactants.

The distinction we make between function/attitude aligns, to some extent, with the broader linguistic/emotional divide. However, we do not impose a strict boundary between these two domains of meaning. Instead, we recognize that certain functional labels may encompass affective or emotional dimensions such as *surprised newsmark* or *expression of excitement*. This occurs when these elements constitute well-defined conversational actions that directly shape the progression of interaction.

Text S5 - Control experiment

As a means of validation of our functional analysis (see Methods Section “Cluster function and attitude”), we conducted the following control experiment. Three naive annotators, with no prior exposure to the clustering solution, were tasked with annotating 80 IUs based on their perceived function. The 80 IUs were randomly selected from the sample of 20 manually-analyzed clusters, provided they were associated with a function in the original analysis. Following a short training session, the annotators were instructed to listen to each example within its wide conversational context and select the most appropriate label from a list including all functions appearing in the sample of 80 IUs ($n=25$). The results indicate an average agreement rate of $71\pm3\%$ with our original analysis (see Table S7). This is a typical rate of successful agreement in such tasks (20). Moreover, disagreements between annotators were often regarding similar functions. To illustrate this, we report an agreement rate of $80\pm2\%$ with our original analysis with respect to function categories (see Table S8). As Table S8 shows, a function category may encompass variations of a single function (e.g., *continuer* vs. *enthusiastic continuer*, or *expression of surprise* vs. *expression of mild surprise*), or functions that share a broader common denominator, such as *strong assessment*, *strong/dramatic statement*, and *surprising statement*, all of which include the conveyance of non-trivial information or opinion.

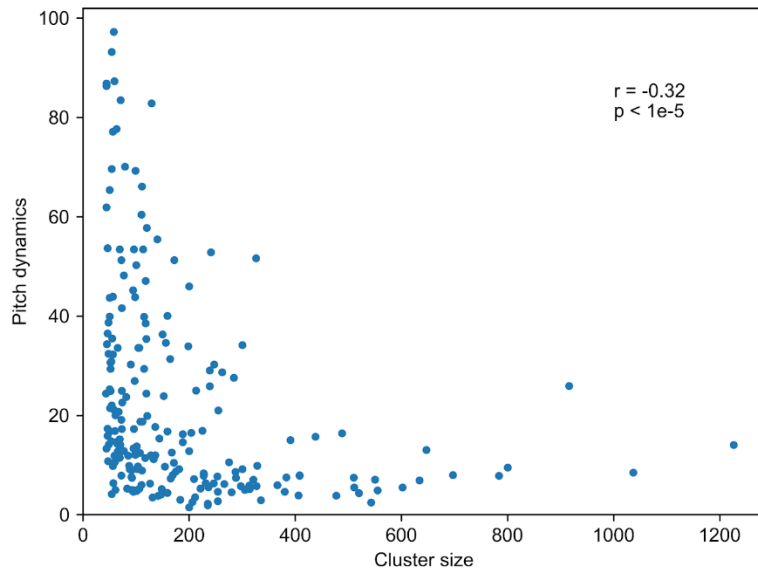


Fig. S1. The correlation between cluster size and a measure of pitch dynamics (see Methods). We find that larger clusters are generally less prosodically marked (21). r - Pearson's correlation.

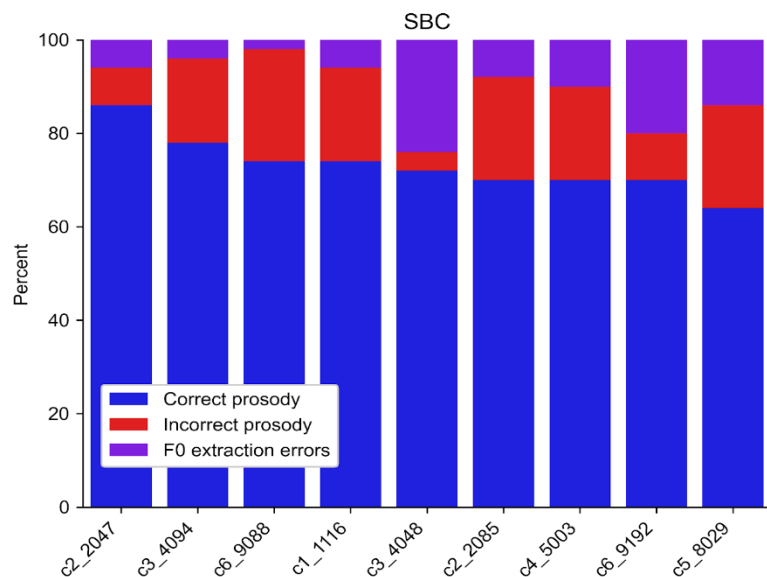


Fig. S2. The distribution of three categories – “correct prosody” (IUs match the prosodic pattern of the cluster), “incorrect prosody” (IUs do not match the prosodic pattern of the cluster), “F0 extraction error” (IUs assigned to the cluster due to an F0 extraction error) – within 9 clusters (n=1695 IUs, SBC dataset).

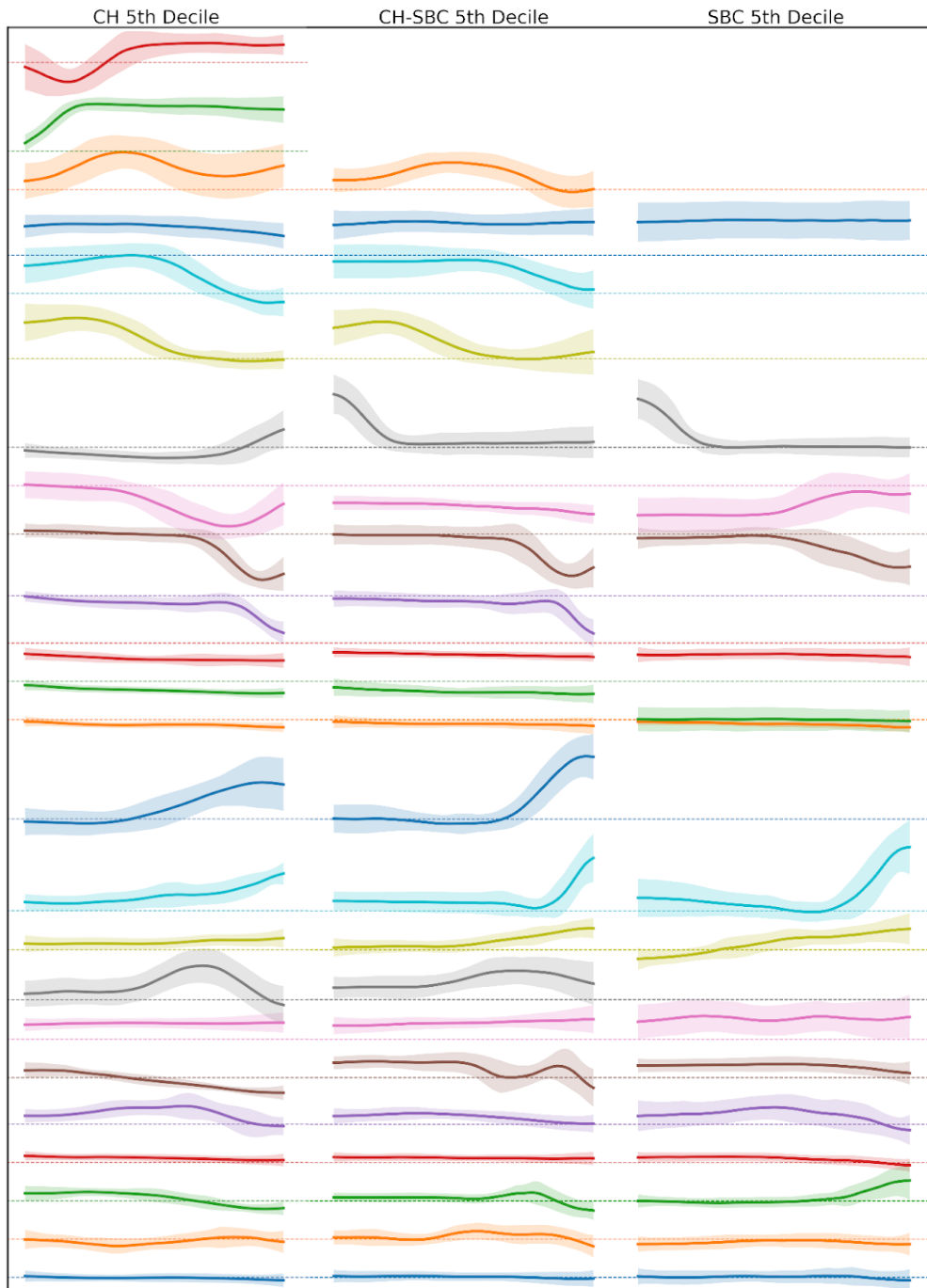


Fig. S3. Mean (\pm standard deviation) pitch contour for each cluster from the 5th decile of the CH (left), SBC (right), and combined analyses (center). Clusters are positioned next to and colored to match the cluster most similar to them (by Euclidean distance). Pitch contours are speaker-normalized, and for each cluster a horizontal dotted line with matching color represents the median of the speaker voice. “Repeated cycle” clusters are not shown.

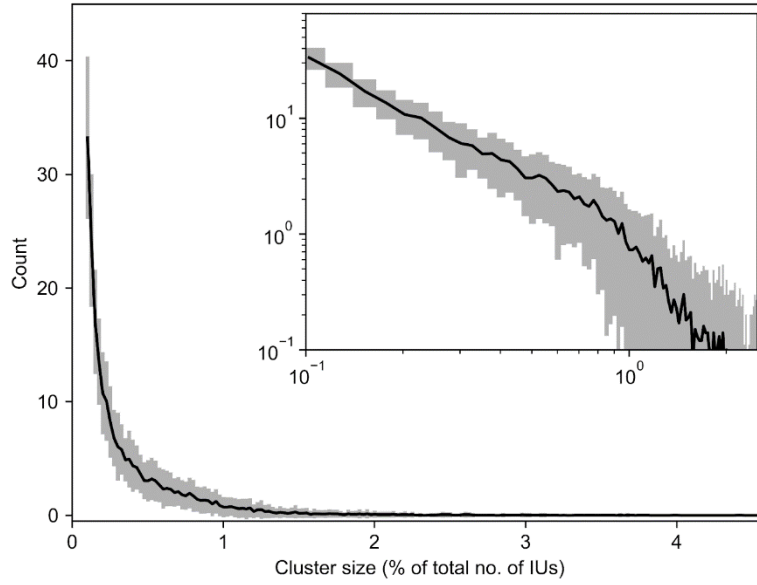


Fig. S4. The distribution of the sizes of clusters in the CH dataset (as a percentage of the total number of IUs in the dataset). Mean (\pm standard deviation) shown for 100 clustering runs with different initial randomization seeds. Same data with log-log axes is shown in the inset. The mean cluster size was $0.39\% \pm 0.003\%$ (\pm standard error).

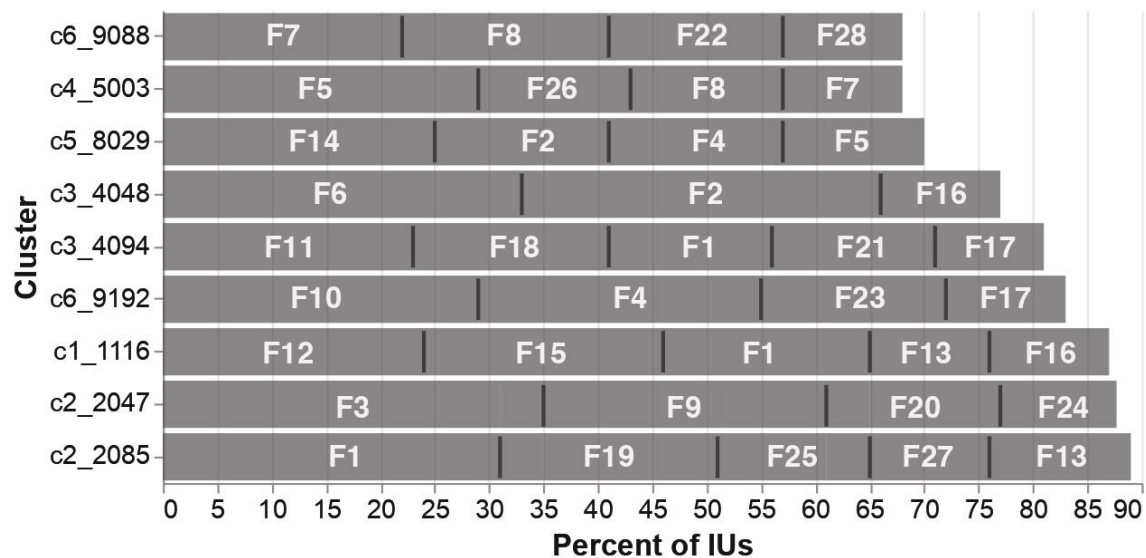


Fig. S5. Summary of the functional analysis of 9 clusters from the SBC dataset. Each bar represents a cluster, with the x-axis indicating the fraction of IUs exhibiting recurring functions. The stacked bars represent the frequencies of the functions in the cluster. Function numbers correspond to Table S4.

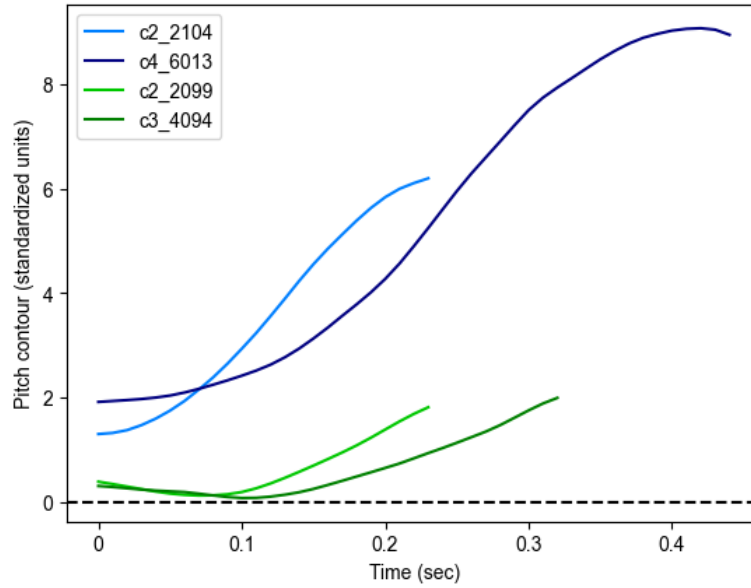


Fig. S6. Similarity of form and function. Clusters c2_2104 and c4_6013 are characterized by a high-range and large pitch rise. Both these prosodic patterns function, inter alia, as *surprised newsmark* and convey the attitude *intrigued-surprised*. Likewise, c2_2099 and c3_4094 are characterized by a mid-range small pitch rise. Both these prosodic patterns function, inter alia, as *continuer*, and convey attitudes which are low in emotive involvement (*passive-disengaged* and *acquiescent*, respectively). The dashed horizontal line at 0 represents the speaker's median pitch.

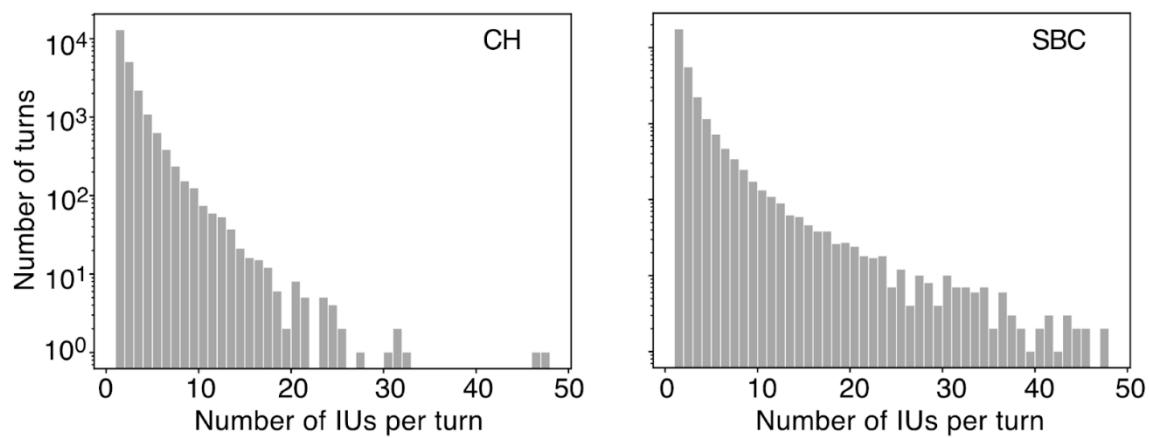


Fig. S7. Turn length distributions in terms of number of IUs, within the CH and the SBC datasets (semi-log). A turn is a single speaker's contribution, delimited by the speech of others.


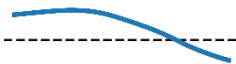

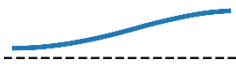
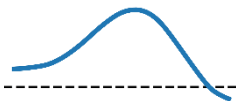
| Cluster | Contour | Description | British School “tone” system (O’Connor & Arnold, 1973) | ToBI annotation system (Beckman et al. 2005) |
|---------|---|---|---|---|
| c1_1013 |  | mid-range pitch fall | <i>Low Fall</i> (<i>The Low Drop</i>) | L* L-L% |
| c3_4000 |  | mid-high-range small pitch rise followed by a large fall | <i>High Fall</i> (<i>The High Drop</i>) | H+L* L-L% |
| c2_2099 |  | mid-range small pitch rise | <i>Low Rise</i> (<i>The Take-off</i>) | L* H-H% |
| c2_2104 |  | high-range large pitch rise | <i>High Rise</i> (<i>The high Bounce</i>) | L+H* H-H% |
| c4_6095 |  | high-range large rise-fall pitch movement | <i>Rise-fall</i> (<i>The Jackknife</i>) | %H H+L* L- L% |

Fig. S8. The internal structure of the pitch contours of the prosodic patterns described in Fig. 2b-c translated into the British school “tone” system (5) and the ToBI annotation system (11, 12). The dashed horizontal line represents the speaker’s median pitch.

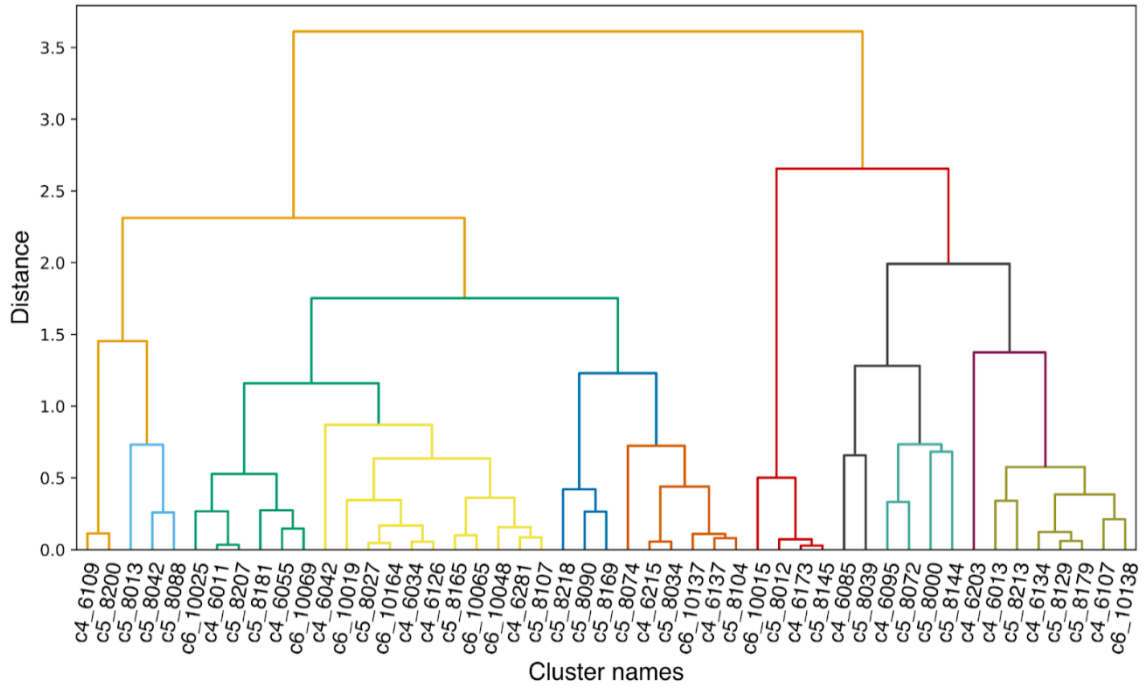


Fig. S9. Hierarchical grouping of cluster pitch contours. Our analysis does not rely on the internal structure of IUs to determine the inventory of pitch contours. However, the centroid pitch contours exhibited similarities in the internal structure and in the prosodic register (the relative height of an entire contour). This figure shows the hierarchical clustering (in real space) of pitch contours from three adjacent deciles (4,5, and 6) of the CH dataset. To do this, each centroid pitch contour was interpolated to the mean size of all the pitch contours that were analyzed from the three deciles - 62 sample points. Distances were calculated as a linear combination of (i) the correlation between the pitch contours allowing for a maximal shift of 30% along the time axis; and (ii) the difference between the registers in semitones. Weights between 0.4-0.6 for the two criteria produced similar dendrograms. While the specific grouping is influenced by the choice of distance metric, several metrics yielded broadly consistent dendrograms. The figure shows that pitch contours can be grouped into “similar” subsets based on their internal structure and register. The groups that correspond to this dendrogram are explicitly plotted in Fig. S10 (below).

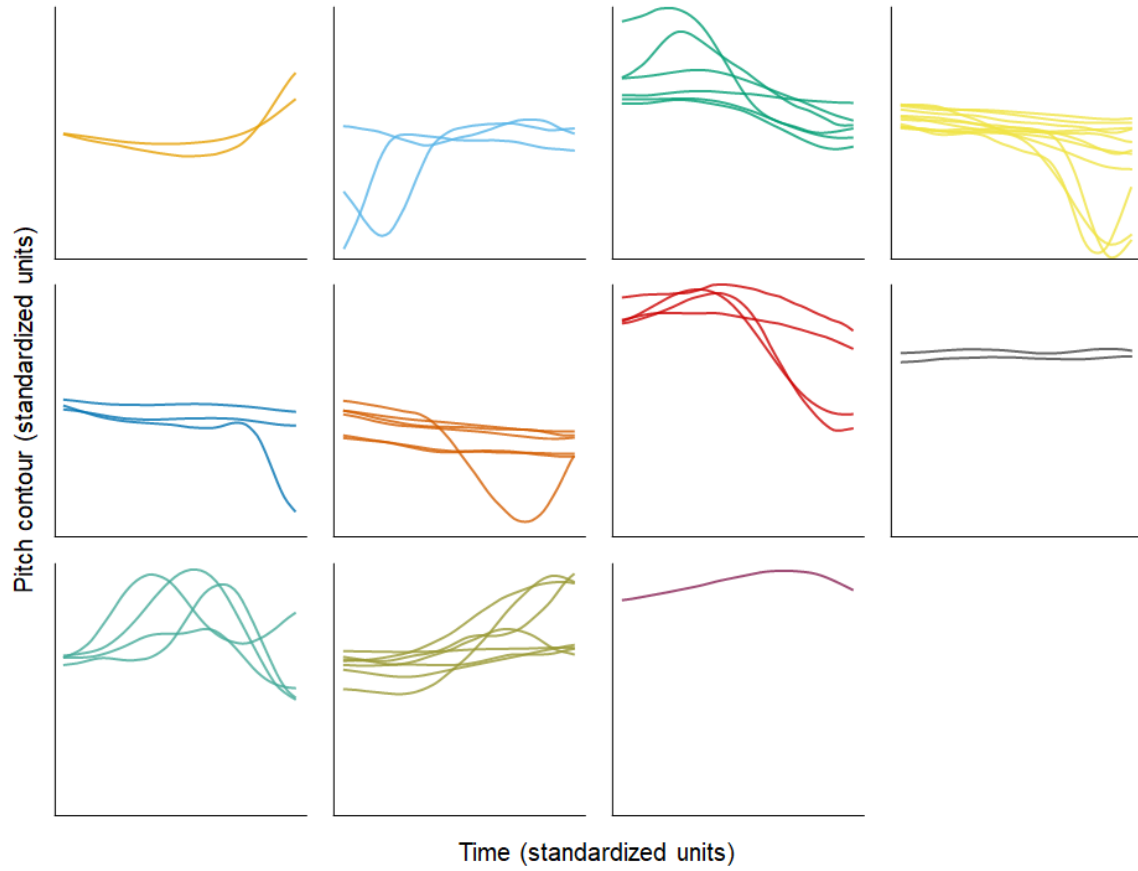


Fig. S10. Groups of pitch contours corresponding to the hierarchical clustering described in Fig. S9 above, obtained with a cutoff distance of 1.15. All panels use the same x-axis scale (62 sample time-points) and the same y-axis scale (pitch in semitones between the minimum -11.9 and the maximum 10.8, where zero was set to the median pitch of the speaker). The colors in the different panels correspond to the colors in Fig. S9.

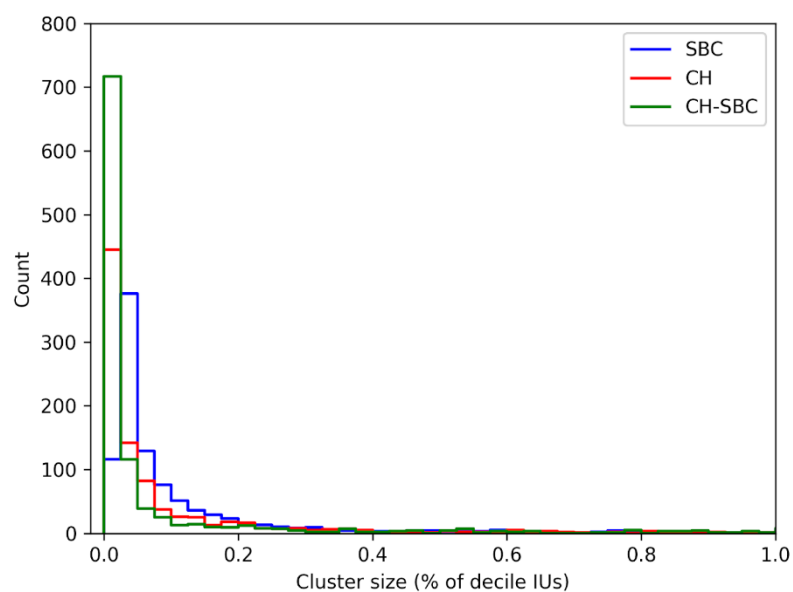


Fig. S11. Distribution of the sizes of clusters below the 1% threshold in the CH dataset (as a percentage of the total number of IUs in the respective decile).

Table S1. The effects of varying the clustering hyperparameters and algorithm on the sizes, number, and silhouette scores of clusters. Each entry represents the mean of 10 clustering runs. We first varied the `n_components` hyperparameter of the GMM clustering algorithm. We then varied the size of the latent space used in the AE network. Third, we looked at the effect of using a different activation function in the AE network. Since the sigmoid activation function generated a clustering result with better silhouette scores, but a larger AE loss after training, we then tried to substantially increase the number of AE training epochs to get to a loss value comparable to that of our original SELU activation function. Finally, we examined the effect of using a different method of clustering altogether - KMeans clustering based on the Euclidean distance between pitch contours (after interpolation and smoothing as in the primary analysis, see Methods). Note that for the sigmoid and KMeans analyses only one decile was processed (decile 5) to reduce processing time. The sigmoid and KMeans clustering results were examined manually as detailed in Table S2 (marked by “Evaluated”). Runs with hyperparameter values equal to the ones used in the primary run are marked with “*”. Rows marked with “#” represent the same data. “Repeated cycle” clusters were ignored. (mean \pm standard deviation)

| Clustering Algorithm | Activation Function | GMM n_components | Latent Dim. | Training Epochs | Mean Cluster Size | Mean Cluster Number | Mean Sil. Score | Mean AE Loss | Deciles Processed | |
|----------------------|---------------------|--------------------------|-------------|-----------------|-------------------|---------------------|-----------------|----------------|-------------------|-----------|
| AE + GMM | SELU | 30 | 8 | 1k | 192 \pm 10 | 215 \pm 12 | 0.06 \pm 0.01 | 1.9 \pm 0.03 | All | |
| AE + GMM | SELU | 300 | 8 | 1k | 202 \pm 23 | 160 \pm 12 | 0.10 \pm 0.01 | 1.9 \pm 0.03 | All | * # |
| AE + GMM | SELU | 3000 | 8 | 1k | 266 \pm 7 | 153 \pm 3 | 0.07 \pm 0.01 | 1.9 \pm 0.03 | All | |
| AE + GMM | SELU | 300 | 4 | 1k | 236 \pm 16 | 181 \pm 11 | 0.16 \pm 0.01 | 3.2 \pm 0.03 | All | |
| AE + GMM | SELU | 300 | 8 | 1k | 202 \pm 23 | 160 \pm 12 | 0.10 \pm 0.01 | 1.9 \pm 0.03 | All | * # |
| AE + GMM | SELU | 300 | 16 | 1k | 186 \pm 12 | 118 \pm 6 | 0.18 \pm 0.01 | 1.4 \pm 0.04 | All | |
| AE + GMM | SELU | 300 | 8 | 1k | 202 \pm 23 | 160 \pm 12 | 0.10 \pm 0.01 | 1.9 \pm 0.03 | All | * # |
| AE + GMM | Sigmoid | 300 | 8 | 1k | 379 \pm 19 | 113 \pm 6 | 0.29 \pm 0.02 | 5.7 \pm 0.19 | All | |
| AE + GMM | SELU | 300 | 8 | 1k | 248 \pm 39 | 13 \pm 2 | 0.09 \pm 0.02 | 1.4 \pm 0.05 | 5th only | * |
| AE + GMM | Sigmoid | 300 | 8 | 1k | 386 \pm 68 | 11 \pm 2 | 0.26 \pm 0.03 | 4.9 \pm 0.08 | 5th only | Evaluated |
| AE + GMM | Sigmoid | 300 | 8 | 100k | 349 \pm 40 | 11 \pm 1 | 0.54 \pm 0.06 | 0.9 \pm 0.05 | 5th only | Evaluated |
| | | KMeans n_clusters | | | | | | | | |
| KMeans | | 30 | | | 180 \pm 9 | 22 \pm 1 | 0.07 \pm 0.01 | | 5th only | Evaluated |
| KMeans | | 300 | | | 64 \pm 3 | 21 \pm 2 | 0.09 \pm 0.01 | | 5th only | |

Table S2. Manual evaluation of the percentage of “correct prosody” IUs (see Methods) in a sub-sample of clusters from several runs with differing hyperparameters, as specified in Table S1 (clusters taken from a single run of each hyperparameter set). We can see that runs using the sigmoid activation function generated inferior clustering results. For the sigmoid activation function and for KMeans clustering, only clusters from decile 5 were examined. “Repeated cycle” clusters were ignored.

| Clustering | Mean % IUs with “correct prosody” | N clusters analyzed | Deciles processed |
|---------------------|-----------------------------------|---------------------|-------------------|
| SELU (primary run) | 70% | 20 | All |
| SELU (primary run) | 70.5% | 2 | 5th only |
| Sigmoid 1k epochs | 53.7% | 6 | 5th only |
| Sigmoid 100k epochs | 50.3% | 6 | 5th only |
| KMeans 30 | 65% | 6 | 5th only |

Table S3. Labels of single-cluster function resulted from manual analysis of 20 clusters from the CH dataset. For each function, the table includes an example, the cluster/s exhibiting the function, and the function relative frequency. (asterisks indicate functions which are by definition interactional, responsive or inviting a response).

| Function | Example | Ex. identifier (file_sec.) | Cluster/s | % |
|--|--|----------------------------|---|------|
| 1. yes-no question* | <i>have you heard from Ron and Doris</i> | 4065_890.91 | c7_12198, c8_15047, c9_16023, c10_18072 | 8.6% |
| 2. continuer* | <i>m-hm</i> | 6217_438.77 | c2_2099, c3_4094 | 6.0% |
| 3. strong/dramatic statement | <i>believe it or not he's an English teacher now</i> | 4792_327.57 | c7_12198, c8_15047, c9_16023, c10_18133 | 4.9% |
| 4. strong/weighty agreement* | <i>of course</i> | 6217_581.59 | c3_4000, c4_6095, c5_8000, | 4.6% |
| 5. complaint | <i>called her yesterday and there was no answer</i> | 4629_206.74 | c9_16023, c10_18072, c10_18133 | 4.4% |
| 6. contrary to previous | <i>no that was when she</i> | 4887_254.58 | c4_6095, c6_10069, c9_16023 | 4.0% |
| 7. affirmation/agreement* | <i>yeah</i> | 5713_693.89 | c1_1013 | 3.8% |
| 8. transition to predication | <i>and then we'll</i> | 4431_499.63 | c6_10065, c10_18133 | 3.8% |
| 9. contrastive statement | <i>it's not dangerous it's dangerous if you're losing weight</i> | 4844_709.91 | c5_8000, c9_16029 | 3.6% |
| 10. expression of excitement* | <i>wow</i> | 4072_50.64 | c4_6095, c6_10069 | 3.5% |
| 11. list item | <i>I like the music</i> | 4522_514.56 | c7_12002, c8_15047, c8_15234 | 3.5% |
| 12. surprising statement | <i>she's actually going to college to be an opera singer</i> | 4838_836.31 | c9_16029, c10_18072 | 2.9% |
| 13. expression of surprise* | <i>oh</i> | 5273_437.83 | c1_1090, c5_8000 | 2.6% |
| 14. strong assessment | <i>it's very cultural</i> | 5273_554.39 | c3_4000, c6_10069 | 2.6% |
| 15. surprised newsmark* | <i>did it</i> | 6252_300.72 | c2_2104, c4_6013 | 2.4% |
| 16. enthusiastic continuer* | <i>m-hm</i> | 4580_335.72 | c2_2104, c4_6013 | 2.2% |
| 17. critical statement | <i>but see it's just like her the last minute be too lazy</i> | 0638_778.1 | c10_18133 | 2.0% |
| 18. expression of empathy* | <i>I know I know</i> | 4556_473.66 | c5_8104, c8_15234 | 2.0% |
| 19. animated newsmark* | <i>you're kidding</i> | 5713_478.8 | c4_6095 | 1.8% |
| 20. basis for dramatic statement | <i>by the time you've been pregnant for</i> | 5373_636.86 | c8_15047, c9_16029 | 1.8% |
| 21. reported speech/behavior | <i>well do you think this is pajamas</i> | 4926_1091.2 | c7_12002 | 1.8% |
| 22. weighty acknowledgement* | <i>oh</i> | 4431_48.8 | c3_4000 | 1.8% |
| 23. conclusive remark | <i>it was so official</i> | 6045_378.68 | c6_10065, c8_15234 | 1.5% |
| 24. initiating dispreferred response* | <i>well</i> | 6045_230.24 | c2_2104, c4_6013 | 1.5% |
| 25. hesitant transition to predication | <i>one is um</i> | 4808_848.08 | c7_12002 | 1.5% |
| 26. newsmark* | <i>yeah</i> | 4875_592.26 | c1_1090, c2_2099 | 1.5% |
| 27. recycled move | <i>you're not into that</i> | 4595_412.68 | c6_10065 | 1.5% |
| 28. repair initiation* | <i>huh</i> | 4580_723.61 | c1_1090 | 1.5% |
| 29. topic/frame shift | <i>like Clair</i> | 6474_256.14 | c5_8000 | 1.5% |
| 30. agreeing 2nd assessment* | <i>that's too weird</i> | 4822_785.81 | c5_8104 | 1.3% |
| 31. filler/placeholder | <i>and um</i> | 4822_796.87 | c5_8104 | 1.3% |
| 32. logical conjunction | <i>so</i> | 4431_125.96 | c1_1013 | 1.3% |
| 33. astonished repair initiation* | <i>what</i> | 6861_854.43 | c2_2104 | 1.1% |
| 34. surprised question* | <i>in Madrid</i> | 4157_493.32 | c4_6013 | 1.1% |

| | | | | |
|----------------------------------|--|--------------|--------------------|------|
| 35. uptalk | <i>I looked into a masters program here</i> | 4807_612.8 | c10_18072 | 1.1% |
| 36. acknowledgement* | <i>good</i> | 5046_266.92 | c1_1013 | 0.9% |
| 37. expression of mild surprise* | <i>oh</i> | 5532_525.52 | c2_2099 | 0.9% |
| 38. intensification | <i>especially if it's just the few days before</i> | 4112_247.31 | c10_18072 | 0.9% |
| 39. presenting reported speech | <i>in his BMW and he's like</i> | 4838_468.54 | c8_15234, c8_15047 | 0.9% |
| 40. rough estimation | <i>like fifteen thousand or whatever here</i> | 4093_1529.38 | c10_18133 | 0.9% |
| 41. weighty logical conjunction | <i>because</i> | 5872_831.67 | c3_4000 | 0.9% |
| 42. contrastive question* | <i>would you rather have a doctor who tells you not to eat</i> | 4753_821.91 | c9_16029 | 0.7% |
| 43. parenthesis | <i>just writing this down so I</i> | 4245_496.15 | c8_15234 | 0.7% |
| 44. repeat | <i>Danny</i> | 6314_593.62 | c3_4000 | 0.7% |
| 45. reluctant agreement* | <i>maybe</i> | 6100_1331.91 | c3_4094 | 0.5% |

Table S4. Labels of single-cluster function resulted from manual analysis of 9 clusters from the SBC dataset. For each function, the table includes the cluster/s exhibiting the function and the function relative frequency. Gray shading indicates 9 functions not identified in the sample analyzed from the CH dataset (see Table S3).

| Function | Cluster/s | % |
|-------------------------------------|---------------------------|------|
| 1. initiating dispreferred response | c1_1116, c2_2085, c3_4094 | 9.2% |
| 2. basis for dramatic statement | c3_4048, c5_8029 | 6.5% |
| 3. pointing at mutual knowledge | c2_2047 | 5.7% |
| 4. reported speech/behavior | c5_8029, c6_9192 | 5.4% |
| 5. transition to predication | c5_5003, c5_8029 | 5.4% |
| 6. contrary to previous | c3_4048 | 4.6% |
| 7. list item | c4_5003, c6_9088 | 4.6% |
| 8. unenthusiastic statement | c4_5003, c6_9088 | 4.6% |
| 9. affirmation/agreement | c2_2047 | 4.2% |
| 10. surprising statement | c6_9192 | 3.8% |
| 11. expression of excitement | c3_4094 | 3.4% |
| 12. logical conjunction | c1_1116 | 3.4% |
| 13. acknowledgement | c1_1116, c2_2085 | 3.1% |
| 14. contrastive statement | c5_8029 | 3.1% |
| 15. expression of negative emotion | c1_1116 | 3.1% |
| 16. feedback elicitation | c1_1116, c3_4048 | 3.1% |
| 17. weighty logical conjunction | c3_4094, c6_9192 | 3.1% |
| 18. basis for strong assessment | c3_4094 | 2.7% |
| 19. hesitant turn initiation | c2_2085 | 2.7% |
| 20. retraction | c2_2047 | 2.7% |
| 21. complaint | c3_4094 | 2.3% |
| 22. non-climatic narrative episode | c6_9088 | 2.3% |
| 23. yes-no question | c6_9192 | 2.3% |
| 24. filler/placeholder | c2_2047 | 1.9% |
| 25. protest | c2_2085 | 1.9% |
| 26. reluctant agreement | c4_5003 | 1.9% |
| 27. expression of surprise | c2_2085 | 1.5% |
| 28. insignificant detail | c6_9088 | 1.5% |

Table S5. Labels of single-cluster attitude resulting from manual analysis.

| Attitude | Cluster |
|--------------------------|-------------------|
| Enthusiastic | c4_6095 |
| puzzled-misinformed | c1_1090 |
| unenthusiastic | c1_1013 |
| passive-disengaged | c2_2099 |
| intrigued-surprised | c2_2104, c4_6013 |
| decisive-authoritative | c3_4000 |
| Acquiescent | c3_4094 |
| aroused-opinionated | c5_8000, c6_10069 |
| routine-mundane | c5_8104, c6_10065 |
| unknowing-interested | c7_12198 |
| Calm | c8_15234 |
| dramatic-engaged | c9_16023 |
| contrary-unexpected | c9_16029 |
| dissatisfied-inquisitive | c10_18072 |
| knowledgeable | c10_18133 |

Table S6. Labels of cluster-pair functions resulted from manual analysis, for 17 cluster-pairs exhibiting $\geq 50\%$ functional uniformity.

| Function | Example (1 st IU 2 nd IU) | Identifier (file_timestamp) | cluster pair | Occur. | Functional uniformity |
|---|---|--------------------------------|-------------------------|--------|--------------------------|
| Additional explanation/elaboration | <i>she's worried about him because Michael has about three kinds of pneumonia going through him right now</i> | 6045_ 489.36 | c10_18091- c10_18053 | 5 | 100% |
| Strong agreement | <i>right exactly</i> | 4157_ 640.91 | c3_4014- c5_8107 | 5 | 80% |
| Reporting habitual behavior or continuous state | <i>I usually I just come home from work and like stay home</i> | 0638_ 239.45 | c9_16027- c6_10137 | 12 | 66.7% |
| Stating a noteworthy fact | <i>the last month I had to see him once a week</i> | 5373_ 552.81 | c6_10137- c7_12110 | 6 | 66.7% |
| Sharing future plans | <i>and I haven't been to Texas in a while so I thought that would work out and</i> | 4660_ 347.37 | c8_15040- c9_16027 | 6 | 66.7% |
| Recycled move | <i>it's tough it's tough</i> | 4104_ 486.56 | c5_8039- c5_8039 | 6 | 66.7% |
| Two-step punchline | <i>which is that we got married</i> | 6137_ 316.99 | c4_6137- c5_8090 | 13 | 61.5% |
| Indecisive answer | <i>well there could be rats I've never seen any</i> | 4665_ 1135.93 | c6_10025- c5_8207 | 5 | 60% |
| Describing undesired scenario | <i>unless I just decide to give up and just take what I've got</i> | 4624_ 642.86 | c9_17139- c5_8104 | 5 | 60% |
| Topic shift | <i>good good for you so you've been leading the life of Riley huh</i> | 4431_ 72.17 | c7_12007- c10_18091 | 5 | 60% |
| Repeating a bottom line | <i>yeah I think about that too</i> | 4721_ 319.56 | c1_1090- c6_10048 | 5 | 60% |
| Assessment + support | <i>the bread is wonderful like everywhere you go there's pita bread and</i> | 4967_ 431.71 | c9_17006- c10_19001 | 5 | 60% |
| Two-step description | <i>see if we can get a one way and then buy another one way going back</i> | 4104_ 551.23 | c8_15234- c9_16026 | 5 | 60% |
| Two-step negative stance taking | <i>and about people coming and seeing me and blah blah blah</i> | 5788_ 131.97 | c9_17006- c6_11001 | 5 | 60% |
| Responsive positive assessment | <i>oh isn't that good oh</i> | 4628_ 262.71 | c2_2124- c6_10069 | 5 | 60% |
| | <i>oh that's good</i> | 4157_ 391.15 | c1_1042- c6_10069 | 6 | 50% |
| Follow-up question | <i>oh when does school start</i> | 0638_ 269.77 | c1_1003- c5_8179 | 6 | 50% |

Table S7. Agreement rate of annotators in control experiment with our original analysis, with respect to exact functions and to function categories (see Table S8).

| | Annotator 1 | Annotator 2 | Annotator 3 | Average |
|--------------------------|-------------|-------------|-------------|---------|
| Function | 76.3% | 72.5% | 63.8% | 70.8% |
| Function category | 80.0% | 83.8% | 76.3% | 80.0% |

Table S8. List of functions labels appearing in the control experiment and their categories.

| Function | Category |
|-----------------------------|---|
| affirmation/agreement | Positive/affiliative responses |
| strong/weighty agreement | |
| agreeing 2nd assessment | |
| weighty acknowledgement | Indications for the reception of new information |
| expression of surprise | |
| expression of mild surprise | |
| expression of empathy | expression of empathy |
| continuer | Go-ahead responses within a long turn of another |
| enthusiastic continuer | |
| contrary to previous | contrastive moves |
| contrastive statement | |
| complaint | Expressions of negative sentiment/valence |
| critical statement | |
| reported speech/behavior | reported speech/behavior |
| filler/placeholder | Conversational "glue", in transition to a point |
| transition to predication | |
| list item | Repetition of linguistic material |
| recycled move | |
| strong assessment | Assertion of non-trivial information/opinion |
| strong/dramatic statement | |
| surprising statement | |
| yes-no question | Question-like moves, anticipating a specific response |
| uptalk | |
| surprised newsmark | |
| repair initiation | |

Table S9. The SWBD-DAMSL set of DA labels (22), compared to the set of labels used in the current study and to the ISO 24617-2 set (23).

| SWBD-DAMSL | Example | Current study (Function # in Table S3) | ISO 24617-2 (Function # in Table S10) |
|----------------------------------|--|--|---|
| 1. 3rd-party-talk | <i>My goodness, Diane, get down from there.</i> | NA | NA |
| 2. Abandoned/Uninterpretable | <i>So, -/</i> | NA | 39 |
| 3. Action-directive | <i>Why don't you go first</i> | NA | 45 |
| 4. Affirmative non-yes answers | <i>It is.</i> | 4, 7 | 1, 3 |
| 5. Agreement/Accept | <i>That's exactly it.</i> | 4 | 1, 12 |
| 6. Apology | <i>I'm sorry.</i> | NA | 5 |
| 7. Appreciation | <i>I can imagine.</i> | 18, 30 | NA |
| 8. Backchannel/Acknowledge | <i>Uh-huh.</i> | 2, 16, 22, 36 | 8 |
| 9. Backchannel-Question | <i>Is that right?</i> | 15, 19, 26 | NA |
| 10. Collaborative Completion | <i>Who aren't contributing.</i> | NA | 9 |
| 11. Conventional-closing | <i>Well, it's been nice talking to you.</i> | NA | 21, 22 |
| 12. Conventional-opening | <i>How are you?</i> | NA | NA |
| 13. Declarative Wh-Question | <i>You are what kind of buff?</i> | 28, 33 | NA |
| 14. Declarative Yes-No-Question | <i>So you can afford to get a house?</i> | 1, 34 | 32 |
| 15. Dispreferred answers | <i>Well, not so much that.</i> | 6, 24 | 17, 18, 28, 38, 47 |
| 16. Downplayer | <i>That's all right.</i> | NA | NA |
| 17. Hedge | <i>I don't know if I'm making any sense or not.</i> | NA | NA |
| 18. Hold before answer/agreement | <i>I'm drawing a blank.</i> | 31 | 44, 54 |
| 19. Maybe/Accept-part | <i>Something like that</i> | 45 | 56 |
| 20. Negative non-no answers | <i>Uh, not a whole lot.</i> | 6, 24 | 17, 18, 28, 38, 47 |
| 21. No answers | <i>No.</i> | 6, 24 | 17, 18 |
| 22. Non-verbal | <i><Laughter>, <Throat clearing></i> | NA | NA |
| 23. Offers, Options & Commits | <i>I'll have to check that out</i> | NA | 26 |
| 24. Open-Question | <i>How about you?</i> | NA | NA |
| 25. Opinion | <i>I think it's great</i> | 14 | 10 |
| 26. Or-Clause | <i>or is it more of a company?</i> | NA | 33 |
| 27. Other | <i>Well give me a break, you know.</i> | NA | NA |
| 28. Other answers | <i>I don't know</i> | NA | NA |
| 29. Quotation | <i>You can't be pregnant and have cats</i> | 21 | NA |
| 30. Reject | <i>Well, no</i> | 6, 24 | 17, 18, 28, 38, 47 |
| 31. Repeat-phrase | <i>Oh, fajitas</i> | 27, 44 | NA |
| 32. Response Acknowledgment | <i>Oh, okay.</i> | 22, 36 | NA |
| 33. Rhetorical-Questions | <i>Who would steal a newspaper?</i> | 42 | NA |
| 34. Self-talk | <i>What's the word I'm looking for</i> | 43 | NA |
| 35. Signal-non-understanding | <i>Excuse me?</i> | 28, 33 | 7 |
| 36. Statement | <i>Me, I'm in the legal department.</i> | 3, 9, 12, 17 | NA |
| 37. Summarize/reformulate | <i>Oh, you mean you switched schools for the kids.</i> | 23 | NA |
| 38. Tag-Question | <i>Right?</i> | 1 | 20 |
| 39. Thanking | <i>Hey thanks a lot</i> | NA | 48 |
| 40. Wh-Question | <i>What did you wear to work today?</i> | NA | 35 |
| 41. Yes answers | <i>Yes.</i> | 7 | NA |
| 42. Yes-No-Question | <i>Do you have to have any special training?</i> | 1 | 34 |
| Overlap (total) | | 25 (60%) | 23 (55%) |

Table S10. The ISO 24617-2 set of DA labels (23), compared to the set of labels used in the current study and to the SWBD-DAMSL set (22).

| ISO 24617-2 | Example | Current study (Function # in Table S3) | SWBD-DAMSL (Function # in Table S9) |
|------------------------------|---|--|---|
| 1. agreement | <i>Exactly</i> | 4 | 5 |
| 2. alloNegative | <i>No no no no no</i> | 6 | 30 |
| 3. alloPositive | <i>Correct</i> | 4, 7 | 4 |
| 4. answer | <i>send error document ready</i> | NA | NA |
| 5. apology | <i>sorry, pick up the oranges</i> | NA | 6 |
| 6. Apology-accept | <i>No problem</i> | NA | NA |
| 7. autoNegative | <i>I beg you pardon</i> | 28, 33 | 35 |
| 8. autoPositive | <i>Uh-huh</i> | 2, 16, 22, 36 | 8 |
| 9. completion | <i>get to Corning</i> | NA | 10 |
| 10. compliment | <i>You look great</i> | 14 | 25 |
| 11. conditional | <i>I'm afraid we don't have time, unless you do it very quickly</i> | NA | NA |
| 12. confirm | <i>Indeed</i> | 4 | 5 |
| 13. congratulation | <i>congratulations!</i> | NA | NA |
| 14. contact-Check | <i>Hello?</i> | NA | NA |
| 15. contact-Indication | <i>Oh hi</i> | NA | NA |
| 16. correctMisspeaking | <i>to pick up the oranges</i> | 6 | NA |
| 17. disagreement | <i>uh... no</i> | 6, 24 | 15, 30 |
| 18. disconfirm | <i>no</i> | 24 | 21 |
| 19. Expression-sympathy | <i>I'm sorry to hear that</i> | 18 | NA |
| 20. feedbackElicitation | <i>Okay?</i> | 1 | 38 |
| 21. Goodbye-init | <i>Bye bye, see you later</i> | NA | 11 |
| 22. Goodbye-return | <i>Bye bye, see you</i> | NA | 11 |
| 23. Greeting-init | <i>Good morning</i> | NA | NA |
| 24. Greeting-return | <i>Good morning</i> | NA | NA |
| 25. instruct | <i>Go right round until you get to just above that</i> | NA | NA |
| 26. offer | <i>I will look that up for you</i> | NA | 23 |
| 27. Offer-accept | <i>Yes please</i> | 4, 7 | 5 |
| 28. Offer-decline | <i>No thank you</i> | 24 | 15, 30 |
| 29. opening | <i>Okay</i> | 32, 41 | NA |
| 30. pausing | <i>Just a moment</i> | 31 | 18 |
| 31. promise | <i>Shall I begin?</i> | NA | NA |
| 32. Question-check | <i>The meeting starts at ten, right?</i> | 1 | 14, 38 |
| 33. Question-choice | <i>Should the telephone cable go in the telephone line slot or in the external line slot?</i> | NA | 26 |
| 34. Question-propositional | <i>Does the meeting start at ten?</i> | 1 | 42 |
| 35. Question-set | <i>What time does the meeting start?</i> | NA | 40 |
| 36. request | <i>Please turn to page five</i> | NA | NA |
| 37. Request-accept | <i>Sure</i> | 4, 7 | 5 |
| 38. Request-decline | <i>Not now.</i> | 24 | 15, 20, 30 |
| 39. retraction | <i>then we're going to g--</i> | NA | 2 |
| 40. Self Introduction-init | <i>Schiphol Information</i> | NA | NA |
| 41. Self Introduction-return | <i>Good morning, this is De Bruin in Arnhem</i> | NA | NA |
| 42. selfCorrection | <i>then we're going to g-- ... turn straight back</i> | NA | NA |
| 43. selfError | <i>yes oh sorry no...</i> | NA | NA |
| 44. stalling | <i>Let me see...</i> | 31 | 18 |
| 45. suggest | <i>Let's wait for the speaker to finish</i> | NA | 3 |
| 46. Suggest-accept | <i>Let's do so</i> | 4, 7 | 5 |
| 47. Suggest-decline | <i>I'd rather not</i> | 24 | 15, 20, 30 |
| 48. thanking | <i>Thanks a lot</i> | NA | 39 |
| 49. Thanking-accept | <i>Don't mention it</i> | NA | NA |
| 50. topicShift | <i>Something else</i> | 29 | NA |
| 51. turnAccept | <i>OK, let me see</i> | NA | NA |
| 52. turnAssign | <i>Craig?</i> | NA | NA |
| 53. turnGrab | <i>Hold on</i> | NA | NA |
| 54. turnKeep | <i>Uh</i> | 31 | 18 |
| 55. turnTake | <i>Uh...</i> | 31 | NA |

| | | | |
|-------------------|---|----------|----------|
| 56. uncertain | <i>That might be a good idea</i> | 45 | 19 |
| 57. unconditional | <i>I'll come tomorrow, no matter what</i> | 3 | NA |
| Overlap (total) | | 28 (49%) | 32 (56%) |

Four supplementary audio files are available separately and via links provided in the main text:

Audio S1. Accompanies Figure 2b and presents the eight IUs consisting only of the word ‘yeah’, taken from four different clusters and exhibiting different prosodic form-function relations.

Audio S2. Accompanies Figure 2c and presents a representative sample of IUs from cluster c4_6095 with varying text, accomplishing three distinct functions and sharing one attitude – “enthusiastic”.

Audio S3. Accompanies Figure 3b and presents two representative instances of cluster pair c9_16027-c6_10137.

Audio S4. Accompanies Figure 3c and presents two representative instances of cluster pair c1_1042-c6_10069.

SI References

1. J. J. Gumperz, *Discourse Strategies* (Cambridge University Press, 1982).
2. P. Auer, "Introduction: John Gumperz' approach to contextualization" in *The Contextualization of Language*, P. Auer, A. Di Luzio, Eds. (John Benjamins Publishing Company, 1992), pp. 1–38.
3. M. Selting, "Prosody as an activity-type distinctive cue in conversation: the case of so-called 'astonished' questions in repair initiation" in *Prosody in Conversation*, E. Couper-Kuhlen, M. Selting, Eds. (Cambridge University Press, 1996), pp. 231–271.
4. A. Cruttenden, *Intonation* (Cambridge University Press, 1997).
5. J. D. O'Connor, G. F. Arnold, *Intonation of colloquial English* (Longman, London, 1973).
6. M. A. K. Halliday, *Intonation and grammar in British English* (De Gruyter, 1967).
7. R. Kingdon, *The groundwork of English intonation* (Longman, 1958).
8. J. Pierrehumbert, "The Phonology and Phonetics of English Intonation," MIT. (1980).
9. M. E. Beckman, J. B. Pierrehumbert, Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255–309 (1986).
10. J. Pierrehumbert, J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse" in *Intentions in Communication*, P. R. Cohen, J. Morgan, M. E. Pollack, Eds. (MIT Press, 1990), pp. 271–311.
11. K. Silverman, *et al.*, ToBI: A Standard for Labelling English Prosody in *The 1992 International Conference on Spoken Language Processing*, (ISCA Archive, 1992).
12. M. E. Beckman, J. Hirschberg, S. Shattuck-Hufnagel, "The Original ToBi System and the Evolution of the ToBi Framework" in *Prosodic Typology*, (Oxford University Press Oxford, 2005), pp. 9–54.
13. D. R. Ladd, *Intonational Phonology*, 2nd Ed. (Cambridge University Press, 2008).
14. J. R. Firth, "A Synopsis of Linguistic Theory, 1930-55" in *Special Volume of the Philological Society*, (Blackwell, 1957), pp. 1–31.
15. J. R. Firth, "Sounds and prosodies" in *Prosodic Analysis*, F. R. Palmer, Ed. (Oxford University Press, 1970), pp. 1–26.
16. M. A. K. Halliday, *An introduction to functional grammar* (Edward Arnold, 1985).
17. J. Local, G. Walker, Methodological Imperatives for Investigating the Phonetic Organization and Phonological Structures of Spontaneous Speech. *Phonetica* 62, 120–130 (2005).
18. E. Couper-Kuhlen, M. Selting, *Prosody in Conversation* (Cambridge University Press, 1996).
19. J. Kelly, J. Local, *Doing phonology: observing, recording, interpreting* (Manchester University Press, 1989).
20. N. Duran, S. Battle, J. Smith, Inter-annotator Agreement Using the Conversation Analysis Modelling Schema, for Dialogue. *Commun Methods Meas* 16, 182–214 (2022).
21. T. Yoshimura, S. Hayamizu, H. Ohmura, K. Tanaka, Pitch pattern clustering of user utterances in human-machine dialogue in *Proceeding of Fourth International Conference on Spoken Language Processing*, (IEEE, 1996), pp. 837–840.
22. A. Stolcke, *et al.*, Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics* 26, 339–373 (2000).
23. H. Bunt, V. Petukhova, A. Malchanau, A. Fang, K. Wijnhoven, The DialogBank: dialogues with interoperable annotations. *Lang Resour Eval* 53, 213–249 (2019).