

Prosody

Fernanda Ferreira¹

¹University of California, Davis

MIT Press

Published on: Apr 02, 2026

URL: <https://oecs.mit.edu/pub/1w4cqquc>

License: [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License \(CC-BY-NC-ND 4.0\)](#)

Prosody refers to the rhythmic and tonal properties of language that transcend individual words or sounds. Core prosodic features include variations in duration (the shortening and stretching of the syllables of a word), stress levels (volume or loudness, often associated with emphasis), and fundamental frequency or pitch (the “tune” of an utterance). The definition of prosody may also be extended to include hesitation pauses, which are periods of silence when a speaker may be planning the next chunk or attempting to retrieve a word. More controversially, filler expressions such as the English “uh” and “um” are sometimes also treated as an aspect of prosody, particularly because they tend to occur in the same environments as silent pauses and may play critical roles in maintaining fluency. Although discussions of prosody traditionally focus on speech, prosody is also a feature of sign languages, in which cues such as larger and longer signs often mark the edges of prosodic units.

History

It has been known for decades that spoken language is organized into prosodic constituents ([Chomsky & Halle, 1968](#); [Ladd, 1996](#); [Lehiste, 1970](#); [Nespor & Vogel, 1986](#)) [see [Language](#)]. For example, if a person were to say, “Whenever I drink wine, I get a headache,” the word “wine” would be elongated and somewhat emphasized, the fundamental frequency of that word would fall and then quickly rise, and a pause might follow. As a result, the listener would perceive the utterance as two intonational phrases separated by a prosodic boundary. Below the intonational phrase are other prosodic constituents, including prosodic phrases, prosodic words, and syllables. In some theories of prosody, this hierarchy of constituents is governed by principles such as the *strict layering hypothesis* ([Selkirk, 1984](#)), which posits that prosodic constituents must nest neatly without overlap so that each constituent is fully housed within higher-level structures. This constraint contrasts with what we see in syntax, in which a clause, for example, may be inserted inside a smaller constituent such as a noun phrase.

The intonational component of prosody refers to variations in pitch, which significantly influence the meaning and interpretation of spoken utterances ([Pierrehumbert & Hirschberg, 1990](#)). Intonational variations may signal grammatical structures; for example, in the previous example, “Whenever I drink wine, I get a headache,” the pitch fall–rise pattern that would typically occur on “wine” helps signal that the sequence will be followed by another clause. Intonation also distinguishes statements from questions, highlights important information, and conveys emotional or pragmatic nuances. By modulating pitch contours, speakers guide listeners’ attention, facilitate comprehension, and convey communicative intent beyond lexical content alone. For example, depending on intonation, the question, “Are you talking to me?” may be received as a sincere question or as a threat (as illustrated in a famous scene from the film *Taxi Driver*).

Theories of how prosodic structure relates to linguistic form fall into two main camps: direct and indirect. In *direct theories*, it is assumed that prosodic features such as phrase-final lengthening, pausing, and pitch movements are direct consequences of syntactic structure. Early experimental work seems to support this view.

For instance, studies have shown that the stronger the syntactic boundary on which a word terminates, the greater the amount of lengthening that is observed, and the higher the probability of a pause ([Cooper & Paccia-Cooper, 1980](#)).

In contrast, *indirect theories* posit the existence of a distinctly prosodic constituency that is derived from syntax but is not identical to it ([Selkirk, 1986](#)). In these approaches, the right edge of a syntactic phrase will typically mark the right edge of a prosodic phrase, but additional syntactic phrase brackets do not strengthen the prosodic boundary. In addition, a recent version of the indirect approach derived from *optimality theory* [see [Phonology](#)] assumes syntax–prosody correspondence but treats the alignment as probabilistic, with the result that some prosodic phrasings may end up violating syntactic constituency in the presence of another more compelling constraint. One such constraint is the attempt to balance the size of phrases ([Fodor, 1998](#)), which may result in a prosodic phrasing such as “(Mary walked) (all the way to the center of town).” In this example, a prosodic phrase boundary occurs after the main verb, which does not align with the right edge of a syntactic phrase but does help balance the size of the prosodic constituents.

Core concepts

Prosody is grounded in interactions between linguistic structure and constraints that arise from the architecture of the human cognitive system [see [Neuroscience of Language](#); [Neuroscience of Syntax](#)]. Linguistic representations influence prosodic features—for example, a clause boundary will almost always trigger an intonational phrase boundary. However, other prosodic features are associated with aspects of performance such as planning—for example, a speaker might slow down and pause at the end of a clause to gain time to plan the next stretch of speech. This prosodic marking of linguistic units such as phrases and clauses benefits the comprehender because prosodic features make abstract syntax concrete and perceptible, which in turn has been shown to facilitate comprehension and learning across a variety of languages, including German ([Lamekina & Meyer, 2023](#)), French ([Didirková et al., 2019](#)), Japanese ([Funasaki & Yano, 2025](#)), and English ([Snedeker & Casserly, 2010](#)), among others [see [Psycholinguistics](#)].

In normal spontaneous speech, intonational phrases tend to be short, with estimates ranging between 1.6 seconds ([Inbar et al., 2025](#)) and three seconds ([Inbar et al., 2020](#)) in duration. Similarly, speakers tend to take a breath every three seconds ([Kallay et al., 2019](#)), and naive listeners report that they perceive speech chunks in unrehearsed speech to be about 2.5 seconds in duration ([Vetchinnikova et al., 2023](#)). Moreover, the best cue to the presence of an intonational phrase boundary is phrase-final lengthening ([Biron et al., 2021](#)), which may reflect the speaker’s need to buy time to recover from the cognitive effort of assembling the current chunk and to plan the next. Pause times within an utterance tend to correlate with latencies to initiate utterances ([Ferreira, 1991](#)), suggesting that both measures reflect planning time.

The correspondence between syntactic and prosodic structures assumed in both direct and indirect theories of prosody implies that syntactic boundaries will tend to be prosodically marked, providing listeners with acoustic cues for parsing, and a large body of research shows that listeners can use this information to choose one syntactic parse over another in cases of syntactic ambiguity. For example, in “Frank saw the man with binoculars,” the parse that treats “with binoculars” as an instrument (rather than a modifier of “man”) would be signaled by the lengthening of “man” ([Snedeker & Trueswell, 2003](#)). In addition, investigations using magnetoencephalography have shown that syntactic boundaries that are marked prosodically are more robustly detected by the human brain ([Degano et al., 2024](#)). Studies of children’s acquisition of language have shown that children use prosodic information to infer syntactic constituents ([Juszyk et al., 1992](#)), an idea known as prosodic bootstrapping ([Gervain et al., 2021](#)). Children leverage prosodic cues to learn vocabulary as well as the grammatical structure of the language to which they are exposed ([De Carvalho et al., 2022](#); [Han et al., 2024](#); [Soderstrom et al., 2003](#)) [see [Word Learning](#)].

In silent reading, people also tend to project a prosodic contour onto the text, generating what is referred to as *implicit prosody* ([Fodor, 1998](#))—an internal representation of the prosodic form the utterance would have if it were spoken. The implicit prosody hypothesis proposes that this covert prosody can influence syntactic parsing decisions. Evidence for these effects comes from work linking punctuation, line breaks, and default prosodic phrasing to comprehension measures such as reading time ([Breen, 2014](#); [Hirose, 2003](#)).

Sign languages also make use of prosody [see [Sign Language](#)]. Signers use manual and nonmanual behaviors to systematically mark prosodic constituency and the units of syntactic and prosodic constituents. The edges of intonational phrases feature boundary cues such as final lengthening of signs, changes in sign size, and pauses. These occur alongside nonmanual boundary markers such as head movements and especially eye blinks, which cluster at prosodic phrase boundaries ([Sandler, 2010](#); [Wilbur, 2022](#)). These cues also perform discourse-level semantic functions such as marking information structure, which refers to the division of an utterance into given and new content. Signers reliably perceive these boundary cues and use them to segment the signed stream into prosodic and syntactic constituents ([Brentari et al., 2018](#)).

Questions, controversies, and new developments

Until recently, our understanding of prosody emerged mainly from studies employing artificial utterances designed to test linguistic hypotheses. Typically, the utterances were elicited by asking speakers to read them aloud, and then the acoustic features were analyzed manually. Today, computational tools for automatic parsing and for performing prosodic analyses allow for a new approach in which spontaneously generated language can be analyzed for both its grammatical and acoustic features and their degree of alignment. This shift towards examining prosody in naturalistic contexts broadens the scope of prosody research. It not only reframes how researchers interpret prosodic markers but also highlights the need to consider performance variables when analyzing language structure and processing. This perspective invites broader investigation into phenomena

traditionally dismissed as noise or errors, such as hesitations and fillers, redefining them as functional aspects of spoken language with cognitive relevance.

Broader connections

Prosody sits at the intersection of language and cognition. It guides how people produce and interpret spoken language in real time, shaping everything from syntactic parsing to discourse interpretation and reflecting cognitive processes such as planning, memory, and attention. It also plays a key role in development, helping infants segment speech and bootstrap syntax [see [Language Acquisition](#)]. Beyond language, prosody signals emotion, speaker intent, and social position, linking linguistic structure to broader systems of perception and communication [see [The Interaction Engine](#)]. In short, studying prosody gives cognitive scientists a window into how the mind handles the complexities of language in context and in interaction.

Further reading

- Cole, J. (2015). Prosody in context: A review. *Language, Cognition and Neuroscience*, 30(1-2), 1-31. <https://doi.org/10.1080/23273798.2014.963130>
- Larrouy-Maestri, P., Poeppel, D., & Pell, M. D. (2025). The sound of emotional prosody: Nearly 3 decades of research and future directions. *Perspectives on Psychological Science*, 20(4), 623-638. <https://doi.org/10.1177/17456916231217722>
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, 25(7-9), 905-945. <https://doi.org/10.1080/01690961003589492>

References

- Biron, T., Baum, D., Freche, D., Matalon, N., Ehrmann, N., Weinreb, E., Moses, E., & Biron, D. (2021). Automatic detection of prosodic boundaries in spontaneous speech. *PLoS One*, 16(5), e0250969. <https://doi.org/10.1371/journal.pone.0250969>
- Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8(2), 37-50. <https://doi.org/10.1111/lnc3.12061>
- Brentari, D., Falk, J., Giannakidou, A., Herrmann, A., Volk, E., & Steinbach, M. (2018). Production and comprehension of prosodic markers in sign language imperatives. *Frontiers in Psychology*, 9, 770. <https://doi.org/10.3389/fpsyg.2018.00770>

- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.

↵

- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Harvard University Press.

↵

- De Carvalho, A., Kolberg, L. S., Trueswell, J., & Christophe, A. (2022). Cross-linguistic evidence for the role of phrasal prosody in syntactic and lexical acquisition. *Speech Prosody*, 396-400.

<https://doi.org/10.21437/SpeechProsody.2022-81>

↵

- Degano, G., Donhauser, P. W., Gwilliams, L., Merlo, P., & Golestani, N. (2024). Speech prosody enhances the neural processing of syntax. *Communications Biology*, 7, 748.

<https://doi.org/10.1038/s42003-024-06444-7>

↵

- Didirková, I., Crible, L., & Simon, A. C. (2019). Impact of prosody on the perception and interpretation of discourse relations: Studies on “et” and “alors” in spoken French. *Discourse Processes*, 56(8), 619-642.

<https://doi.org/10.1080/0163853X.2018.1528963>

↵

- Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances.

Journal of Memory and Language, 30(2), 210–233. [https://doi.org/10.1016/0749-596X\(91\)90004-4](https://doi.org/10.1016/0749-596X(91)90004-4)

↵

- Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research*, 27(2), 285-319.

<https://doi.org/10.1023/A:1023258301588>

↵

- Funasaki, N., & Yano, M. (2025). Role of prosody and word order in identifying focus: Evidence from pupillometry. *Language, Cognition and Neuroscience*, 40(1), 23–40.

<https://doi.org/10.1080/23273798.2024.2396962>

↵

- 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 University Press

↵

- Han, M., De Jong, M., De Jong, N.H., & Kager, R. (2024). Relating the prosody of infant-directed speech to children’s vocabulary size. *Journal of Child Language*, 51(1), 217-233.

<https://doi.org/10.1017/S0305000923000041>

↑

- Hirose, Y. (2003). Recycling prosodic boundaries. *Journal of Psycholinguistic Research*, 32(2), 167–195.
<https://doi.org/10.1023/a:1022448308035>

↑

- Inbar, M., Grossman, E., & Landau, A. N. (2020). Sequences of intonation units form a ~1 Hz rhythm. *Scientific Reports*, 10, 15846. <https://doi.org/10.1038/s41598-020-72739-4>

↑

- Inbar, M., Grossman, E., & Landau, A. N. (2025). A universal of speech timing: Intonation units form low-frequency rhythms. *Proceedings of the National Academy of Sciences*, 122(34), e2425166122.
<https://doi.org/10.1073/pnas.2425166122>

↑

- Jusczyk, P. W., Hirsh-Pasek, K., Nelson, D. G., Kennedy, L., & Woodward, A. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24(2), 252–293.
[https://doi.org/10.1016/0010-0285\(92\)90009-Q](https://doi.org/10.1016/0010-0285(92)90009-Q)

↑

- Kallay, J. E., Mayr, U., & Redford, M. A. (2019). Characterizing the coordination of speech production and breathing. *Journal of the Acoustical Society of America*, 144(3 suppl), 1720-1721.
<https://doi.org/10.1121/1.5067628>

↑

- Ladd, D. R. (1996). *Intonational phonology*. Cambridge University Press.

↑

- Lamekina, Y., & Meyer, L. (2023). Entrainment to speech prosody influences subsequent sentence comprehension. *Language, Cognition and Neuroscience*, 38(3), 263-276.
<https://doi.org/10.1080/23273798.2022.2107689>

↑

- Lehiste, I. (1970). *Suprasegmentals*. MIT Press.

↑

- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Foris.

↑

- Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press.

↵

- Sandler, W. (2010). Prosody and syntax in sign languages. *Transactions of the Philological Society*, 108(3), 298–328. <https://doi.org/10.1111/j.1467-968X.2010.01242.x>

↵

- Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. MIT Press.

↵

- Selkirk, E. O. (1986). *On derived domains in sentence phonology*. In M. Aronoff & R. T. Oehrle (Eds.), *Language sound structure: Studies in phonology presented to Morris Halle* (pp. 371–405). MIT Press.

↵

- Snedeker, J., & Casserly, E. (2010). Is it all relative? Effects of prosodic boundaries on the comprehension and production of attachment ambiguities. *Language and Cognitive Processes*, 25(7-9), 1234–1264. <https://doi.org/10.1080/01690960903525499>

↵

- Snedeker, J., & Trueswell, J. C. (2003). Using prosody to avoid ambiguity: Effects of speaker awareness and referential context. *Journal of Memory and Language*, 48(1), 103–130. [https://doi.org/10.1016/S0749-596X\(02\)00519-3](https://doi.org/10.1016/S0749-596X(02)00519-3)

↵

- Soderstrom, M., Seidl, A., Nelson, D. G. K., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49(2), 249–267. [https://doi.org/10.1016/S0749-596X\(03\)00024-X](https://doi.org/10.1016/S0749-596X(03)00024-X)

↵

- Vetchinnikova, S., Konina, A., Williams, N., Mikusová, N., & Mauranen, A. (2023). Chunking up speech in real-time: Linguistic predictors and cognitive constraints. *Language and Cognition*, 15(3), 453–479. <https://doi.org/10.1017/langcog.2023.8>

↵

- Wilbur, R.B. (2022). Prosody in sign languages. *Hrvatska revija za rehabilitacijska istraživanja*, 58, 143–174. <https://doi.org/10.31299/hrri.58.si.8>

↵