Studies in Second Language Acquisition (2026), 1–20
doi:10.1017/S0272263126101788

DATA REPORT

English Prosody Dataset for investigating English prosody acquisition

Magdalena Kachlicka[1], Kyle Jasmin[2], Ashley Symons[2], Kazuya Saito[3,4], Fred Dick[5] and Adam Tierney[1]

[1]Birkbeck, University of London, School of Psychological Sciences, London, UK; [2]Royal Holloway, University of London, Department of Psychology, Egham, UK; [3]University College London, Institute of Education, Department of Culture Communication and Media, London, UK; [4]Tohoku University, School of International Studies, Sendai, Japan and [5]University College London, Faculty of Brain Sciences, Division of Psychology and Language Sciences, London, UK
Corresponding author: Magdalena Kachlicka; Email: m.kachlicka@bbk.ac.uk

(Received 19 July 2025 UTC; Revised 03 April 2026 UTC; Accepted 11 April 2026 UTC)

Abstract

English prosody conveys grammatical structure, word meaning, and pragmatic information. However, research on second language (L2) acquisition has primarily focused on segmental properties of speech, with relatively less attention to prosody. We present the English Prosody Dataset to facilitate research on the acquisition and use of English prosody. The dataset includes contrastive examples of three prosodic features: phrase boundary, disambiguating syntactic dependencies; contrastive focus, disambiguating key information within a phrase; and lexical stress, disambiguating word sense through syllabic emphasis, all recorded by three female and three male voice actors speaking Standard Southern British English. Researchers can tailor experiments by selecting stimuli isolating individual acoustic dimensions (e.g., pitch and duration) and by adjusting perceptual difficulty using samples with reduced acoustic information. The dataset also provides acoustically unmanipulated recordings and sentence lists that can be used to develop naturalistic comprehension paradigms or to elicit speech.

Keywords: contrastive focus; lexical stress; phrase boundary; prosody

Introduction

Speech prosody is essential to communication, helping distinguish questions from statements, signaling conversational turn-taking, emphasizing new or important information, and marking syntactic boundaries. Through variations in pitch, intensity, duration, and rhythm, speakers convey meaning, emotion, and intent beyond the literal words they use. Despite the recognized importance of prosody in speech

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (http://creativecommons.org/licenses/by-nc-sa/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is used to distribute the re-used or adapted article and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

perception, teaching English prosody is not always incorporated into second language (L2) learning curricula (Lengeris, 2012; for review of trends in L2 speech perception training see: Ingvalson, Ettlinger, & Wong, 2014; for discussion on effectiveness of prosody training see: Chun & Levis, 2020) and can be challenging for English language learners (e.g., Polish, Rojczyk & Porzuczek, 2012; Mandarin Chinese, Yu & Gibbon, 2015; French, Frost & Picavet, 2014; Korean, Shin & Speer, 2012). The relatively lesser emphasis on prosody compared to pronunciation or grammar in teaching English as a foreign language (EFL) may stem from the assumption that prosody only conveys nuanced lexical and phrasal information that is context-specific. By contrast, mispronunciation of individual phonemes can obviously change the intended word's meaning or referent. Nonetheless, while contextual information may sometimes compensate for misinterpreted or incorrectly produced prosody, prosodic mistakes can lead to communication breakdowns. For example, misplacing a pause in a sentence like "Let's eat, Grandma!" versus "Let's eat Grandma!" can mean the difference between a warm invitation and a scenario more akin to 19th-century German fairytales. The relatively limited focus on speech prosody in L2 pedagogy is even more surprising given that very often, intelligible and comprehensible pronunciation is set as a goal for L2 learners. Moreover, accurate prosody is associated with the degree of perceived intelligibility (Kang, Thomson, & Moran, 2020), comprehensibility, and nativelikeness (Trofimovich & Baker, 2006).

There has been continued interest in English prosody in second language acquisition research (e.g., Baek, 2021, 2024; Liu, 2021; Sun, Saito, & Tierney, 2021; Kachlicka et al., 2025; for discussion about the place of suprasegmentals in EFL teaching see: Wang, 2020; for overview of prior work on prosody see: *at the word level*, Jongman & Tremblay, 2020; *at the sentence level*, Trouvain & Braun, 2020), emphasizing a range of difficulties experienced by learners. What makes understanding prosody in a new language particularly challenging is that acoustic patterns conveying prosody are not uniform across languages. In English, word stress placement is not entirely predictable and must be memorized as a part of each word's pronunciation (Cutler, 2015). This unpredictability can cause substantial difficulty for L2 English learners, particularly those whose native languages have fixed-stress patterns such as Polish (primary stress on penultimate syllable; Wierzchowska, 1971; Domahs, Knaus, Orzechowska, & Wiese, 2012), French (final syllable stress; Peperkamp & Dupoux, 2002), Finnish, or Hungarian (initial syllable stress; Peperkamp, Vendelin, & Dupoux, 2010). Acoustic cues used to identify word stress also differ across learners from various language backgrounds (for review see: Gordon & Roettger, 2017). For instance, while vowel quality was shown to be the strongest cue to perceiving lexical stress for all groups, pitch served as the second strongest cue for native speakers of English and Mandarin Chinese. In contrast, Russian speakers tend to disregard pitch and instead rely more heavily on duration and intensity to determine stress (Chrabaszcz, Winn, Lin, & Idsardi, 2014). Korean speakers also used vowel quality as a primary cue, similarly to native speakers, using F0 and intensity to a lesser degree (Lee, 2022).

Similarly at the sentence level, patterns of acoustic prominence in English can signal meaning by bringing information into focus. As a linguistic phenomenon, focus is an integral part of sentences' semantic structure (Rooth, 1992) and serves to emphasize new or important information (identificational vs. contrastive focus, Kiss, 1998; but see Gussenhoven, 2008 for a more in-depth discussion describing five additional types of focus in English, namely: corrective, counter-presuppositional, definitional, contingency, and reactivating). However, not all languages utilize prominence in the same

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

way (e.g., Russian vs. English, Ionin, Luchkina, & Goldshtein, 2024; French vs. English, Vander Klok, Goad, & Wagner, 2018; Mandarin vs. English, Chikako et al., 2018; Taiwanese Mandarin vs. English, Tseng & Su, 2014), and some languages such as Northern Sotho from Bantu language family lack prosodic focus-marking (Zerbian, 2015). These differences add another level of difficulty to L2 learning and contribute to varied language learning outcomes. For example, Spanish learners of English have trouble perceiving L2 sentence focus, possibly due to the rigidity of the phrasal prominence in their L1 (Lee, Shin, & Martinez Garcia, 2019), while Mandarin Chinese speakers rely more on F0 contour to perceive the prominence (Chen, 2025) drawing on the importance of F0 in their L1. Cantonese Chinese speakers also use pitch as a primary cue in production: they produce emphasized words with consistently higher F0, even though they can reliably use duration and intensity cues to signal sentence emphasis (although they use intensity less than L1 English speakers, Ng & Chen, 2011).

Cross-linguistic differences are also evident in how acoustic cues signal phrase boundaries (e.g., *perception*, English vs. Japanese, Pinter, Mizuguchi, & Tateishi, 2014; English vs. Mandarin Chinese, Zhang et al., 2018; Kuang, Chan, & Rhee, 2022; Fang, Li, Yu, Schwieter, & Wang, 2025; English vs. French, Gilbert, Lee, Wolpert, & Baum, 2023; *production*, English vs. Korean, Baek, 2021; English vs. German, O'Brien, Jackson, & Gardner, 2014). While phonological phrase-final lengthening is often considered a reliable cue for segmentation, even in unfamiliar languages, its interpretation is not universal. For example, Ordin and colleagues (2017) demonstrated that German and Spanish-Basque native speakers benefited from word-final lengthening when segmenting an artificial language, whereas Italian native speakers showed greater sensitivity to penultimate syllable lengthening. Conversely, antepenultimate lengthening impeded boundary detection by Spanish and Italian speakers, suggesting that prosodic cues are interpreted according to language-specific stress patterns and may hinder segmentation when they conflict with listeners' native prosodic expectations. Indeed, when comparing the perception of English phrase boundaries by native speakers of English and Mandarin Chinese, clear differences in cue use emerged (Zhang, 2012). English listeners relied predominantly on pauses, using pitch and lengthening of pre-boundary word to a lesser extent, whereas Mandarin Chinese listeners relied more heavily on pitch than on the other two cues. In another study comparing phrase boundary production of native English speakers and Korean learners of English, English L1 speakers used pauses, elevated pitch, and intensity, while Korean learners of English relied mainly on pre-boundary lengthening and pause (Baek, 2021).

Given the relatively small selection of tools designed specifically for prosody assessments (for review see: Kalathouttukaren, Purdy, & Ballard, 2015), existing studies on prosody acquisition often rely on datasets that comprise large corpora of annotated spontaneous speech, recorded audio-video material, or short story readings (e.g., Aix-MARSEC, Auran, Bouzon, & Hirst, 2004; LeaP: The Learning the Prosody of a foreign Language, Edalatishams, 2017; Speak & Improve Corpus 2025, Knill, Nicholls, Gales, Qian, & Stroinski, 2024). Such resources allow researchers to investigate authentic prosodic patterns in real-life scenarios. Since prosody is embedded within the context of longer utterances, they can also explore questions about the overall L2 fluency or L1 transfer effects. Other studies use selections of more experimentally controlled stimuli, for example, lists of words (nouns vs. verbs, Kelly, 1988), pseudowords (Chrabaszcz et al., 2014) or both (Yu & Andruski, 2009), in which the placement of lexical stress varies. Other types of stimuli include lists of contrastive early and late phrase boundary sentences

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

(Frazier & Rayner, 1982—partially used and complemented with new sentences by Peter, McArthur, & Crain, 2014; Kjelgaard & Speer, 1999—partially used and complemented with new sentences by Jasmin, Dick, & Tierney, 2020). While these resources are invaluable for studying English prosody and can be found in supplementary materials and associated repositories, none of these datasets was designed to systematically capture multiple prosodic contrasts across multiple talkers. They were also not designed to conduct controlled comparisons of specific prosodic features (large corpora, while potentially containing enough examples, require careful and labor-intensive selection to achieve the balanced, controlled sets needed), investigate learners' ability to generalize prosodic patterns (systematically varied stimuli absent in natural corpora), or cue use for prosodic contrasts (natural corpora do not control for acoustic variability across tokens).

Current study

Here we present and make publicly available the English Prosody Dataset that builds on the existing Multidimensional Battery of Prosody Perception (MBOPP; Jasmin et al., 2020). Our database includes recordings of three aspects of prosody: *phrase boundary* examples capturing the differences in phrase transitions in compound sentences, *contrastive focus* examples capturing the differences in emphasis placed on relevant information within a phrase, and *lexical stress* examples capturing the differences in syllable prominence within a word. Using speech morphing, we isolated individual acoustic dimensions (pitch and duration) across all stimuli, allowing researchers to measure their relative importance in prosody perception. We chose pitch, as it is an important cue to lexical stress (used as a primary cue by L2 learners, e.g., Japanese, Beckman, 1986; Dutch, Tremblay et al., 2021; not used or used less than other cues, e.g., Russian, Chrabaszcz et al., 2014; Korean, Lee, 2022) and contrastive focus in English (e.g., Mandarin Chinese, Chen, 2025), and duration as it is an important cue to phrase boundaries (used as a primary cue by L2 learners, e.g., Korean, Baek, 2021; Mandarin Chinese, Zhang, 2012; German and Spanish-Basque, Ordin et al., 2017); importantly, these two dimensions are orthogonal. This contrast is particularly useful for examining English prosody acquisition in tonal language speakers who tend to overweigh pitch information in L2 speech perception and production (Zhang, Nissen, & Francis, 2008) and have trouble disengaging attention from pitch even when explicitly asked to do so (Jasmin, Sun, & Tierney, 2021). It must be noted, however, that while this selection may be useful for comparing performance across some language groups, for comparing others, morphing stimuli with alternative acoustic manipulations may be needed. The stimuli sample the acoustic space continuously along these dimensions, enabling adjustment of task difficulty as needed.

To enhance generalizability, we provide recordings from six voice actors (three male and three female). In addition, we provide original, unmanipulated recordings and lists of sentences illustrating contrastive examples of phrase boundary placement, contrastive focus, and lexical stress. These materials can be used to develop naturalistic experimental paradigms, L2 prosody training (e.g., high-variability phonetic training), or to elicit production data from L2 and L1 learners. The recordings from the English Prosody Database can also serve as a baseline performance for evaluating English prosody acquisition, as they provide a standard or typical example of how prosody is used by native British English speakers. The provided materials are suitable for testing

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

both prosody perception and production in English across participants from diverse language backgrounds and proficiency levels. The dataset is available in open access and can be downloaded from: https://osf.io/pkrh8/.

The English Prosody Dataset

Materials

The lists of phrase boundary and linguistic focus sentences were taken from the Multidimensional Battery of Prosody Perception[1] (MBOPP; Jasmin et al., 2020) dataset. We added a new set of sentences to capture lexical stress contrasts. The final list of stimuli includes 42 pairs of sentences for *phrase boundary*, 47 for *contrastive focus*, and 50 for *lexical stress.* The sentences were arranged into pairs forming target prosodic contrasts (see Table 1).

Phrase boundary

The *phrase boundary* stimuli are 42 pairs of short sentences with a subordinate clause followed by a main clause. Both carrier sentences in each pair were identical in written form up until the end of the second noun, with the subsequent words forming either an early or late closure of the sentence. In the first type of sentence ("early closure"), the verb in the subordinate clause was intransitive, so that the noun that followed became

Table 1. Examples of Prosodic Contrasts for Phrase Boundaries, Contrastive Focus, and Lexical Stress. Capitalization indicates contrastive stress or emphasis, and commas mark the position of phrase boundaries. Presented here are the lexically identical fragments illustrating the prosodic contrast (target contrast) and the natural context in which they were embedded (carrier sentence)

	Token A	Token B
Phrase boundary	***Early closure***	***Late closure***
Target contrast	Because Mike paid, the bill	Because Mike paid the bill
Carrier sentence	Because Mike paid, the bill was smaller.	Because Mike paid the bill, it was smaller.
Contrastive focus	***Early focus***	***Late focus***
Target contrast	READ books	Read BOOKS
Carrier sentence	Mary likes to READ books, but she doesn't like to WRITE books.	Mary likes to read BOOKS, but she doesn't like to read MAGAZINES.
Lexical stress	***1st syllable stress***	***2nd syllable stress***
Target contrast	CONtract	conTRACT
Carrier sentence	All employees have a written CONtract of employment.	The heart muscles conTRACT to move the blood.

[1]While the recordings of phrase boundary and contrastive focus produced by one voice actor are indeed the same as in MBOPP, we include additional resources related to those recordings, namely: original recordings of all contrasts, original recordings of full sentences, time-annotated morphing substrates that can be used to generate morphed WAV files along any selected acoustic dimensions, and generated audio files morphed along pitch and duration in 5% increments along both dimensions (not just a selection of pitch and duration values, as in MBOPP). Additionally, we provide all of the above recordings and morphs for 5 additional voice actors (6 in total) and include additional lexical stress examples. We also provide the following additional resources: scripts that can be used to generate WAV files morphed along any selected acoustic dimensions (not only pitch and duration) and images with lexical prompts designed specifically for eliciting prosody contrasts in production.

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

the subject of a new clause (e.g., "Because Mike paid, the bill is smaller"). In the second type ("late closure"), the verb was transitive and took the following noun as its object, resulting in the phrase boundary falling later in the sentence ("Because Mike paid the bill, it was smaller"). The recordings used for morphing (i.e., target contrasts) were edited such that they included only the lexically identical portions of the two recordings (i.e., "Because Mike paid the bill" and "Because Mike paid, the bill"), with the pause after the comma removed.

Contrastive focus

The *contrastive focus* stimuli are 47 pairs of compound sentences. Carrier sentences in each pair included two independent clauses separated by a conjunction, forming sentences with early or late contrastive focus. In the first type of the sentence ("early focus") the first word was emphasized (e.g., "Mary likes to READ books, but she doesn't like to WRITE books"), whereas in the second type ("late closure") the second word was emphasized ("Mary likes to read BOOKS, but she doesn't like to read MAGAZINES"). Here also, to extract target contrasts, the recordings were edited such that they included only the first clause of the sentence that contained identical written words but differed in the placement of contrastive focus (i.e., "READ books" and "read BOOKS").

Lexical stress

The *lexical stress* stimuli are 50 pairs of short sentences containing target words with lexical stress placed either on the first or second syllable. Target words were two-syllable words with stress either on the first or second syllable. They were embedded in carrier sentences to capture natural variation in pronunciation. Since the meaning of two contrastive words with lexical stress is always different, it was impossible to create identically sounding sentences. Therefore, pairs of carrier sentences differed in structure and meaning, but were matched in length and included only frequently occurring words and words that are acquired early during L2 learning (all words in carrier sentences were checked against vocabulary profiles; Nation, 2006; Cobb, 2012; Table 2). This makes the sentences easily understandable by L2 English learners at various levels of proficiency. However, researchers using any subset of the stimuli should check the vocabulary profile of their specific selection, as lexical frequency and register may vary across items. The target words (i.e., identical words with stress placed either on the first or second syllable) were extracted from carrier sentences for morphing.

Table 2. Summary of Vocabulary Profiles of Carrier Sentences and Target Words and Phrases. Values represent cumulative % of words per level

	Carrier Sentences			Target Words/Phrases		
Level	Lexical Stress	Contrastive Focus	Phrase Boundary	Lexical Stress	Contrastive Focus	Phrase Boundary
K–1	77.7	99.7	89.7	97.8	100	89.9
K–2	87.7	99.7	95.7	–	–	95.8
K–3	97.6	99.8	96.8	100	–	96.6
K–4	99.9	–	97.4	–	–	97.5
K–5	100	–	98.9	–	–	98.7
≥K–6	–	100	100	–	–	100

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

Table 3. Summary of Talkers' Age, Pitch, and Speech Rates Measured in a Single Test Recording

	Female Voice Actors		Male Voice Actors	
	Values	M	Values	M
Age	24, 33, 43	33.33	25, 36, 46	35.67
Pitch (Hz)	229.51, 248.77, 206.65	228.31	145.18, 111.27, 123.83	126.76
Speech rate (words per minute)	116.24, 108.4, 127.19	117.39	175.62, 136.02, 140.06	150.57

Voice actors

The dataset consists of recordings made by six professional voice actors (three males, three females) speaking with a Standard Southern British English accent. The recordings include a range of talkers of different ages, speaking with different pitch and speech rates (see Table 3).

Speech recordings

Recordings of one male actor were the recordings of phrase and focus reported earlier in the MBOPP battery (Jasmin et al., 2020); the same actor recorded lexical stress stimuli to complement the current dataset. Five additional voice actors recorded stimuli in their home studios following the procedures described in Kachlicka et al. (2025). To ensure that the audio quality of new recordings was comparable with the original recordings from the MBOPP dataset and consistent across talkers, we evaluated the quality of actors' audio setup. Actors first recorded a single test stimulus that was carefully examined by the researcher for audio quality (e.g., background noise, volume, proximity to microphone). After completing this step, actors attended an online training session during which the researcher provided them with detailed instructions as to how they should record the stimuli. They were asked to read the sentences aloud using their natural voice and to make sure that their speech sounded as natural as possible while emphasizing the target contrast. After a short rehearsal, actors were provided with a complete list of stimuli to read and record. All samples were recorded as high-resolution two-channel wav files with a sampling rate of 44.1 kHz and 16-bit audio bit depth.

Pre-processing

Pre-processing involved several steps. First, audio files of full carrier sentences were trimmed to remove any silences at the beginning and end of each recording, ramped with 10-ms on/off linear ramp, and downmixed to mono by averaging the existing channels.

Next, to prepare the audio files for acoustic morphing, an additional step involved extracting target words from the carrier sentences for stress stimuli, as well as extracting lexically identical portions of recordings for linguistic focus and phrase boundary stimuli. Furthermore, for phrase boundary stimuli, silent pauses after phrase breaks were also removed. The extracted targets were ramped with 10-ms on/off linear ramp to avoid acoustic transients. Then, all samples (carrier sentences and target contrasts) were normalized to a target loudness level of −20 dB RMS with the "Match loudness"

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

function in Adobe Audition software (Adobe Inc.). Normalization was implemented to equalize the volume across samples coming from different talkers, not to control their output volume.

Morphing

Morphing was performed with STRAIGHT software (Kawahara & Irino, 2005). During the procedure, each input audio file was decomposed by STRAIGHT into three high-level acoustic features: F0 contour, spectral envelope, and aperiodicity index. The F0 contour information was used to detect voiced and unvoiced parts of recordings. Next, all the features were dynamically matched along the time dimension to create a similarity matrix between the two contrastive recordings. Then, the two recordings were time-aligned by manually inserting anchor points marking identical portions of both recordings in the created similarity matrix. These markers included onsets of words, syllables, and salient phonemes. Since morphing generates audio by interpolating between the two input recordings, precise alignment ensured that morphing would affect only the differing parts of both recordings, that is, the selected prosodic contrasts. The resulting morphing substrates were then used to warp all the features and re-scale the values along a particular dimension to resynthesize the stimuli. The synthesized audio files represent intermediate versions of the two input recordings, gradually shifting in pitch and duration from one to the other (e.g., from early to late boundary). An experimenter listened to the resulting audio files to ensure they were of good quality and sounded as natural as possible, repeating the process and adjusting anchor points until the audio quality of all synthesized samples was satisfactory. Stimuli that did not meet quality criteria were not included in the dataset. It should be noted that some degree of acoustic distortion is an inevitable consequence of acoustic manipulation. However, all included stimuli were judged to be clearly intelligible and speech-like.

STRAIGHT allows morphing along several dimensions: aperiodicity, spectrum, frequency, time (duration), and F0 (pitch). In our dataset, only duration and pitch information were manipulated. In English, pitch and duration covary such that, for example, longer duration and higher pitch signal emphasis or lexical stress, whereas shorter duration and lower F0 values are associated with lack of emphasis or stress. The other dimensions available in STRAIGHT (i.e., aperiodicity, spectrum, and frequency) were set at 50% such that morphs contained equal amounts of information from the two recordings. However, researchers can synthesize stimuli along any of these dimensions, if needed, using the provided morphing substrates.

For all samples, we generated recordings in which the pitch and duration excursions in the diagnostic segments of the syllables or words were shifted from 0% (pitch or duration information only conveying one interpretation) to 100% (pitch/duration only conveying the other interpretation) in 5% increments. For example, for duration manipulations, the resulting morphs of words "COM-pound" (stress placed fully on the first syllable) and "com-POUND" (stress placed fully on the second syllable) correspond to the two duration endpoints. A stimulus with 0% duration represents one end of the continuum (i.e., the word "COM-pound," with first stressed syllable longer than the second) and a stimulus with 100% duration represents the other extreme of the continuum (i.e., the word "com-POUND," with second stressed syllable longer than the first). The stimuli with all the remaining values of duration are perceptually the in-between versions that change smoothly from "COM-pound" to "com-POUND" depending on the available information.

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

Table 4. Overview of English Prosody Database Contents. The underscores represent a folder structure such that "audio_files," "sentence_lists," "image_examples," and "tools" are the main folders, and the number of underscores indicates the number of subdirectories (e.g., a single underscore represents one directory down, a double underscore two)

Data Type	Folders and Files	Description
Audio recordings	audio_files	
Target words	_target_words __untransformed ___phrase ___focus ___stress __morphed ___phrase ___focus ___stress	Contains excerpts of target words or phrases for all prosody contrasts; two types of recordings are available: (1) untransformed natural speech and (2) morphed speech with varying amount of pitch and duration information
Carrier sentences	_carrier_sentences __phrase __focus __stress	Contains original recordings of all carrier sentences
Lists of sentences	sentence_lists	Contains lists of prosody contrast sentences
Image examples	image_examples _phrase _focus _stress	Contains images with lexical prompts for eliciting prosody contrast in speech production
Tools	tools _scripts _morphing_substrates __phrase __focus __stress	Contains example scripts and morphing substrates used to generate morphed WAV stimuli

Dataset structure and contents

In the English Prosody Database, we provide the original recordings of full carrier sentences in which target prosody contrasts were embedded. We also include the extracted target words (lexical stress), word pairs (contrastive focus), or contrastive phrases (boundary) in their original untransformed form and synthesized morphs. An overview of dataset contents is presented in Table 4.

The file naming format for carrier sentences and original untransformed targets is as follows: **[voice ID]_[prosodic feature and stimulus number]_[token type]_[stimulus type].wav**. For example:

- "M01_focus01_early_full.wav" and "M01_focus01_late_full.wav" represent two contrastive recordings of carrier sentences with early and late focus recorded by a male voice actor.
- "F01_stress01_early_target.wav" and "F01_stress01_late_target.wav" represent contrastive recordings of extracted target words with lexical stress placed on first and second syllables recorded by a female voice actor.

The morphed stimuli folder contains recordings in which both pitch and duration increased from 0% to 100% in 5% increments. We provide synthesized recordings of all combinations of pitch and duration contents, resulting in a full grid of morphs that can be sampled according to the researchers' needs. The synthesized recordings can be

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

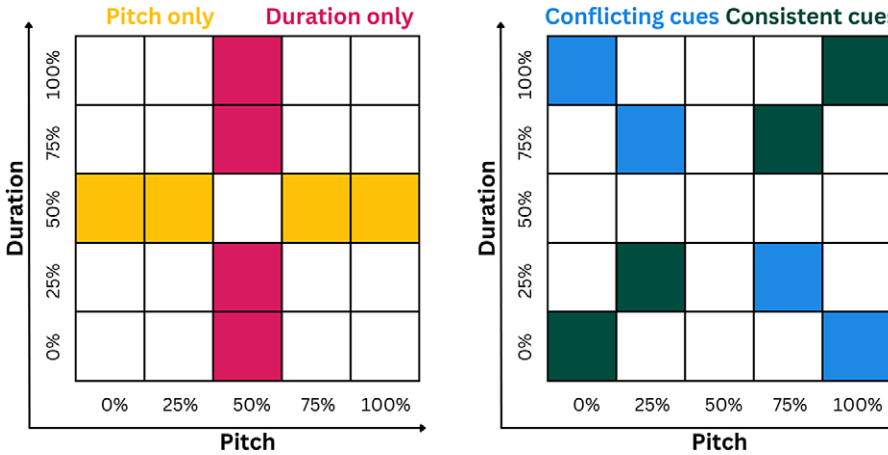


Figure 1. Diagrammatic Representation of the Acoustic Stimulus Space. Stimuli sampled a full 21 x 21 acoustic space across duration and F0 contour in 5% increments. For simplicity, the figure represents only the 25% increments, but morphs with all values are provided. (Left) Acoustic space of stimuli where either pitch or duration is available (i.e., pitch or duration only). (Right) Acoustic space of stimuli where both cues are present, but cue the interpretation of the stimulus in either consistent or conflicting way.

classified into two types depending on their acoustic content. The first type is the recordings where either pitch or duration information is available (i.e., *pitch only* or *duration only*; Figure 1). In the pitch-only stimuli, prosody is conveyed by pitch cues alone which vary from 0% (pitch information coming entirely from the first syllable/early focus/early phrase recordings) to 100% (pitch information coming from the second syllable/late focus/late phrase recordings), while duration cues are set at 50% and are perceptually ambiguous. In the duration-only stimuli, prosody is conveyed only by durational cues varying from 0% to 100%, while pitch cues are set at 50%. The second type are the recordings where both pitch and duration are present, but they cue the interpretation of the stimulus either in a consistent or conflicting fashion (i.e., *conflicting cues* or *consistent cues*; Figure 1).

Additionally, morphs of recordings with consistent cues (e.g., pitch and duration both set to 90%) can be used to manipulate the perceptual difficulty of the contrast, helping to avoid ceiling effects in experimental tasks. Lower percentage values reflect the reduced acoustic difference between stimulus pairs (e.g., first- and second-syllable stress), thereby naturally increasing task difficulty. The lower the percentage, the more perceptually challenging the contrast would be.

The file naming format for carrier sentences and original untransformed targets is as follows: **[voice ID]_[prosodic feature and stimulus number]_[pitch morphing rate]_[duration morphing rate].wav.** For example:

- "M04_focus01_pitch0_duration0.wav" – pitch and duration both cue early contrastive focus interpretation (consistent cues)
- "M04_focus01_pitch20_duration20.wav" – pitch and duration both cue early focus (consistent cues), but the contrast is perceptually more challenging
- "M04_focus01_pitch100_duration100.wav" – pitch and duration both cue late focus (consistent cues)

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

- "M04_focus01_pitch80_duration80.wav" – pitch and duration both cue late focus (consistent cues), but the contrast is more perceptually challenging
- "M04_focus01_pitch100_duration0.wav" – pitch cues late focus and duration cues early focus (conflicting cues)
- "M04_focus01_pitch0_duration100.wav" – pitch cues early focus and duration late focus (conflicting cues)
- "M04_focus01_pitch50_duration100.wav" – pitch is ambiguous, only duration cues late focus
- "M04_focus01_pitch50_duration0.wav" – pitch is ambiguous, only duration cues early focus
- "M04_focus01_pitch100_duration50.wav" – duration is ambiguous, only pitch cues late focus
- "M04_focus01_pitch0_duration50.wav" – duration is ambiguous, only duration cues early focus
- "M04_focus01_pitch50_duration50.wav" – perceptually ambiguous stimulus

Availability

The dataset is available at https://osf.io/pkrh8/ repository and https://sla-speech-tools.com/ website (Mora-Plaza, Saito, Suzukida, Dewaele, & Tierney, 2022).

Discussion

The English Prosody Database has the potential to reinvigorate research on acquisition of English prosody by offering researchers a rich set of stimuli suitable for exploring a wide range of research questions. In the following section, we provide examples of how the stimuli from the English Prosody Database can be used. We then discuss several limitations of the corpus. Finally, we conclude by outlining future directions for the development of the database and prosody research.

Potential applications in L2 and L1 learning research

One of the key strengths of this dataset lies in its adaptability for studying *perceptual cue weighting during prosody perception*. With the provided acoustically manipulated stimuli, researchers can explore how listeners use pitch and duration while perceiving contrastive examples of linguistic focus, lexical stress, and phrase boundaries. Several recent studies have used such stimuli to explore these issues. For example, Jasmin et al. (2021) examined the perception of phrase boundaries using a 5 × 5 stimulus space that varied across pitch and duration. They found that Mandarin speakers relied more heavily on pitch and less on duration when categorizing phrase boundaries ("If Barbara gives up, the ship" vs. "If Barbara gives up a ship") compared to native English and Spanish speakers. The same phrase boundary categorization task has been used in a study exploring cue weighting in Mandarin Chinese L2 learners of English. Petrova and colleagues (2023) investigated how cue weighting strategies shift across L2 learner populations with differing degrees of experience and lengths of residency. By examining how learners adjust their reliance on pitch versus duration, this research contributed to better understanding of how L2 listeners acquire prosodic sensitivity in a new language. The presented dataset can support further investigations into the mechanisms underlying cue weighting during L2 prosody acquisition, such as L1-specific

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

cue-weighting profiles, developmental changes with increased L2 experience, or training-induced cue reweighting. In future research this corpus could also be used to investigate developmental changes in L1 prosodic cue weighting. Although there have been a number of studies of how the weighting of cues to segmental contrasts change with development (Krause, 1982; Nittrouer 2002; Mayo, Scobbie, Hewlett, & Waters, 2003; Mayo & Turk, 2005; Idemaru & Holt, 2013; Nittrouer & Lowenstein, 2015), much less is known about how prosody perception strategies change as children develop.

Researchers can also make use of the dataset's flexibility to *evaluate the precision of prosody perception* and *adjust task difficulty*, if needed. In the study by Sun, Saito, and Tierney (2021), for example, the cues to contrastive linguistic focus were scaled to 60% of their original magnitude to avoid ceiling effects and better capture variability in learners' performance. Another example of difficulty scaling was implemented by Kachlicka et al. (2025) in their perceptual prosody training. In this longitudinal study, learners practiced categorizing prosodic contrasts with real-time feedback over the course of 6 days. The design used acoustic information gradients as a way to control the perceptual difficulty of the stimuli throughout the training. The training began with "easy stimuli," where pitch information was either removed (whispered speech) or made perceptually ambiguous (i.e., set at 50% between the contrastive recordings). This way, the only cue learners could use to determine the category of each stimulus was duration (see "duration only" stimuli in the current dataset). As participants progressed through the training, the size of the duration cues gradually decreased, keeping the training perceptually challenging but still manageable depending on each individual's progress. This training aimed to shift learners' attention from pitch to duration. However, with our dataset, one could just as easily implement a training that focuses on pitch instead. This ability to adaptively modulate the availability of cues, and consequently the difficulty of prosody perception, enables the creation of training paradigms that are suitable for a wide variety of listeners, including individuals with difficulty perceiving prosody. This could include children with a diagnosis of language impairment (Fisher, Plante, Vance, Gerken, & Glattke, 2007; Mundy & Carroll, 2012; Cumming, Wilson, & Goswami, 2015), as well as individuals from L2 backgrounds who tend to struggle to perceive prosodic features in English, such as L1 speakers of fixed stress languages like French and Polish (Peperkamp et al., 2010).

The ability to modify the difficulty of prosody perception tasks can be useful not only to study individuals who struggle to perceive prosodic features but also to avoid ceiling effects when investigating individuals who may be highly proficient at prosody perception. Musical experience, for example, has been linked to better L2 (Kolinsky, Cuvelier, Goetry, Peretz, & Morais, 2009) and L1 (Choi, 2022) English stress perception, but it is unknown whether this advantage extends to other features. L1 tone language speakers may also display an advantage at perception of certain prosodic features in English. Cantonese-English bilinguals, for example, display superior English lexical stress perception relative to English monolinguals (Choi, Tong, & Samuel, 2019), and future work could investigate whether this advantage extends to other prosodic features.

It is also possible to use these stimuli to *introduce unknown accents* into perceptual tasks. One example is the paradigm developed by Jasmin, Tierney, Obasih, and Holt (2023), in which participants were exposed to both familiar and unfamiliar accents to test competing hypotheses about how multidimensional speech categories are processed. Authors hypothesized that if a unified prosodic category is activated during exposure to accented speech, then the inconsistent duration information will lead to a down-weighting of that cue in later tasks. If, however, pitch and duration are evaluated

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

independently, then such inconsistency should have no effect on subsequent cue weighting. The stimuli were sampled from a 7 × 7 acoustic space, with either consistent (canonical) or conflicting (accented) distributions of pitch (F0) and duration. The study showed that listeners rapidly down-weighted duration (a secondary cue) when categorizing accented speech, suggesting that they can track short-term speech regularities and dynamically adjust their perceptual strategies.

An additional advantage of this dataset is its inclusion of recordings by multiple voices and a wide variety of examples, which will *improve the generalizability of experimental findings*. Previous studies of prosodic cue weighting have generally relied on a single stimulus (e.g., "read books" for linguistic focus, Sun et al., 2021) recorded by a single male speaker, which does not capture the natural variability present in everyday speech. Such a design also introduces practical issues, including participant fatigue, as repeated exposure to the same word or phrase for 25 minutes can become quite tiresome and negatively affect performance. In contrast, the current dataset offers a wide range of stimuli recorded by six different speakers, allowing researchers to create more engaging and ecologically valid tasks. This diversity will help sustain participant attention, reduce the likelihood of habituation, and support more generalizable conclusions.

The dataset includes carefully curated materials targeting prosodic features, making it suitable for structured testing of L2 prosody production as well. For example, the original recordings of carrier sentences are an excellent resource for designing *prosody imitation tasks* during which participants first listen to target spoken utterances and then imitate not only the words but also the prosodic intonation. Because the carrier sentences were specifically designed to elicit distinct prosodic contrasts, they provide a controlled and targeted basis for assessing whether learners can perceive and reproduce meaningful prosodic distinctions. These tasks can be used to assess learners' L2 proficiency (Wu, Tio, & Ortega, 2021; for meta-analytic review, see Kostromitina & Plonsky, 2022) and their ability to convey prosody in L2 English in a native-like fashion. Whether elicited imitation performance reflects learners' prosodic knowledge specifically, or if it can tap into automatized grammatical knowledge (Suzuki & DeKeyser, 2015) remains an open question, one this dataset is well positioned to help address.

These stimuli could also be used as targets for production training, although there is evidence that training involving self-imitation rather than imitation of native speakers may be comparatively more effective (Kusz, 2022). The provided recordings can also serve as a benchmark for evaluating learners' performance throughout learning, as they exemplify baseline productions by native English speakers. This can be especially useful in the context of high-level L2 acquisition and identifying which variables predict the attainment of near-nativelike L2 performance. For example, using this dataset, we can precisely compare how closely Chinese listeners (more reliance on pitch) replicate native English listeners' processing of prosody (equal focus on pitch and duration), and examine which individual factors are associated with such near-native processing patterns.

Limitations

Despite its flexibility, there are limitations to this dataset that researchers may wish to consider before use. First, it reflects only Standard Southern British English pronunciation. While this accent is widely studied and often used as a reference accent in linguistic research, it represents just one variety among the many global forms of

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

English. This narrow focus may limit the generalizability of findings to other dialects, such as American, Australian, Canadian, and other regional or national varieties of English, or large English-speaking populations in countries like India or Nigeria, to name a few. These variants of English are not only widely spoken but also highly visible in media, education, and technology. For example, American English is the dominant variety in international film or digital assistants.

With regard to stimuli, it should be acknowledged that although the recordings capture prosodic contrasts within full sentences which constitute their naturalistic context, the sentences themselves are not necessarily natural or realistic (e.g., it is unlikely that one would hear the sentence "If Barbara gives up, the ship will be plundered" in a casual pub conversation, unless interlocutors happen to discuss a strategic navy game). Additionally, many of the sentences are structurally quite awkward, designed primarily to force the contrast of interest, and often lack relatable meaning. While this is not inherently problematic, learners might question the utility or relevance of such sentences, particularly if the purpose of the task is not clearly explained by the researchers. Although efforts were made to make the stimuli accessible to all listeners, the vocabulary and syntactic structures used may not be appropriate for younger or less experienced learners. This is due to both their linguistic complexity and the potential lack of understanding of the concept of prosody in these groups.

While every effort was made to ensure the morphed stimuli were as intelligible and speech-like as possible, it should be acknowledged that some degree of acoustic distortion is an inevitable consequence of acoustic manipulation, and morphed stimuli cannot be expected to be indistinguishable from natural speech. Formal naturalness ratings were not collected as part of the dataset preparation, and researchers with specific naturalness requirements are encouraged to conduct their own ratings on their chosen subset of stimuli prior to use.

Finally, there are a few limitations related to the acoustic features we selected for morphing. Speech is a highly redundant signal (e.g., voicing of a single consonant can be conveyed with as many as 16 different cues; Lisker, 1986). Therefore, including acoustic variations only along pitch and duration limits the range of research questions that can be addressed. However, since we are not the authors of the STRAIGHT morphing software, we cannot share it here (see Belin & Kawahara, 2025 for an updated open-access version of the package). Nevertheless, we have made our scripts and morphing substrates (i.e., time-aligned annotations of contrastive recordings) publicly available. These resources can be used to synthesize new stimuli along other dimensions, if needed.

Future directions

As mentioned earlier, the dataset's current focus on Standard Southern British English limits its generalizability. One particularly relevant issue is the variation in prosodic patterns across different English dialects. Intonation contours and stress placement can vary substantially between dialects and can be a source of difficulty even for native speakers of English. For example, native speakers of Canadian English have been shown to struggle with prosodic cues when listening to British English, demonstrating response patterns more similar to those of L2 speakers (Arnhold et al., 2020). These findings suggest that different accentuation patterns may lead to distinct mental representations, which may not be equally accessible to speakers of various dialects. Differences between native accents can also negatively affect processing speed under

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

challenging conditions (Adank, Evans, Stuart-Smith, & Scott, 2009) and may even alter the perceived meaning of words. A study comparing American and British accents, for example, found that the semantically ambiguous word "bonnet" is interpreted differently depending on the accent in which it is spoken (Cai et al., 2017). Moreover, substantial individual differences exist even among speakers from notionally homogenous language backgrounds (English speakers from London, UK; Peppé, Maxim, & Wells, 2000). Future versions of the database should aim to incorporate multiple dialects, enabling the construction of a more representative corpus.

Existing research suggests that the expression of focus may depend not only on speakers' first language (L1), but also on the specific context in which focus is used. For instance, Mandarin speakers, like English speakers, signal focus by manipulating duration, pitch range, and intensity. However, their use of these cues varies depending on both the type of focus (e.g., greater increase in pitch range, mean, and maximum for signaling new-information) and the discourse-pragmatic context of the conversation (e.g., stronger prosodic effects during negotiation than when reporting a past event; Ip & Cutler, 2016). In another study, speakers of closely related languages like Taiwanese, Taiwan Mandarin, and Beijing Mandarin all raised pitch and intensity of focused words, but only Beijing Mandarin speakers lowered pitch and intensity on post-focus words (Chen, Wang, & Xu, 2009). The unavailability of a preferred cue was also shown to contribute to lower accuracy in focus recognition (Chen et al., 2009). Future versions of our dataset could include examples that would allow to test focus in various contexts.

Finally, while studies investigating cue weighting highlight the role of language-specific patterns in prosodic marking, they also point to a bigger challenge: prosodic cues rarely function in isolation. The realizations of focus, stress, and phrase boundaries emerge from the interplay between phonetic and phonological information. Other sources of variation, such as coarticulation (Repp, 1983), intonation effects (Peng, Lu, & Chatterjee, 2009), and systematic interactions between cues for individual phonemes and speech prosody (de Pijper & Sanderman, 1994; Reinisch, Jesse, & McQueen, 2011) can also influence how prosodic categories are realized and interpreted across languages. These factors should be explored in more detail in future investigations of L2 prosody acquisition.

Data availability statement. All data presented in the manuscript are available via a link to OSF: https://osf.io/pkrh8/.

Declaration of competing interests. Authors declare no conflicts of interest.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 520–529. https://psycnet.apa.org/doi/10.1037/a0013552

Arnhold, A., Porretta, V., Chen, A., Verstegen, S. A. J. M., Mok, I., & Järvikivi, J. (2020). (Mis)understanding your native language: Regional accent impedes processing of information status. *Psychonomic Bulletin & Review*, *27*, 801–808. https://doi.org/10.3758/s13423-020-01731-w

Auran, C., Bouzon, C., & Hirst, D. (2004). The Aix-MARSEC project: An evolutive database of spoken British English. *Speech Prosody (Nara, Japan)*. https://doi.org/10.21437/SpeechProsody.2004-130

Baek, H. (2021). Prosodic disambiguation in first and second language production: English and Korean. *Language and Speech*, *65*(3), 598–624. https://doi.org/10.1177/00238309211042041

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

Baek, H. (2024). Acoustics of the alignment of narrow focus prosody and lexical stress in native and nonnative English productions. *The Journal of the Acoustical Society of America*, *156*, 2340–2350. https://doi.org/10.1121/10.0030469

Beckman, M. E. (1986). *Stress and non-stress accent*. Foris.

Belin, P., & Kawahara, H. (2025). STRAIGHTMORPH: A voice morphing tool for research in voice communication sciences. *Open Research Europe*, *10*(4), 154. https://doi.org/10.12688/openreseurope.18055.2

Cai, Z. G., Gilbert, R. A., Davis, M. H., Gaskell, M. G., Farrar, L., Adler, S., & Rodd, J. M. (2017). Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition. *Cognitive Psychology*, *98*, 73–101. https://doi.org/10.1016/j.cogpsych.2017.08.003

Chen, Y. (2025). *Acoustic cues to prosodic prominence perception for bilingual speakers.* PhD Thesis. The University of Texas at Austin. https://doi.org/10.26153/tsw/62037

Chen, S., Wang, B., & Xu, Y. (2009). *Closely related languages, different ways of realizing focus. Conference of the International Speech Communication Association (INTERSPEECH)*, 1007–1010. https://doi.org/10.21437/INTERSPEECH.2009-298

Chikako, T., Kao, S., Baek, H., Yeung, A. H. L., Hwang, J., & Broselow, E. (2018). Native and non-native speaker processing and production of contrastive focus prosody. *Proceedings of the Linguistic Society of America*, *3*(35), 1–13. https://doi.org/10.3765/plsa.v3i1.4323

Choi, W., Tong, X., & Samuel, A. G. (2019). Better than native: tone language experience enhances English lexical stress discrimination in Cantonese-English bilingual listeners. *Cognition*, *189*, 188–192. https://doi.org/10.1016/j.cognition.2019.04.004

Choi, W. (2022). Towards a native OPERA hypothesis: Musicianship and English stress perception. *Language and Speech*, *65*(3), 697–712. https://awspntest.apa.org/doi/10.1177/00238309211049458

Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research*, *57*, 1468–1479. https://doi.org/10.1044/2014_JSLHR-L-13-0279

Chun, D. M., & Levis, J. M. (2020). Prosody in second language teaching: Methodologies and effectiveness. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 619–630). Oxford University Press.

Cobb, T. (2012). Compleat lexical tutor for data-driven language learning on the web. https://www.lextutor.ca/

Cumming, R., Wilson, A., & Goswami, U. (2015). Basic auditory processing and sensitivity to prosodic structure in children with specific language impairments: a new look at a perceptual hypothesis. *Frontiers in Psychology*, *6*, 972. https://doi.org/10.3389/fpsyg.2015.00972

Cutler, A. (2015). Lexical stress in English pronunciation. In M. Reed & J. M. Lewis (Eds.) *The Handbook of English Pronunciation*. Wiley.

De Pijper, J. R., & Sanderman, A. A. (1994). On the perceptual strength of prosodic boundaries and its relation to suprasegmental cues. *The Journal of the Acoustical Society of America*, *96*(4), 2037–2047. https://doi.org/10.1121/1.410145

Domahs, U., Knaus, J., Orzechowska, P., & Wiese, R. (2012). Stress "deafness" in a language with fixed word stress: An ERP study on Polish. *Frontiers in Psychology*, *3*, 439. https://doi.org/10.3389/fpsyg.2012.00439

Edalatishams, I. (2017). LeaP corpus (Review). In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, (pp. 236–240). Iowa State University.

Fang, L., Li, Z., Yu, K., Schwieter, J. W., & Wang, R. (2025). Hearing tones, missing boundaries: Cross-level selective transfer of prosodic boundaries among Chinese–English learners. *Behavioral Sciences*, *15*(12), 1605. https://doi.org/10.3390/bs15121605

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*, 178–210.

Fisher, J., Plante, E., Vance, R., Gerken, L., & Glattke, T. J. (2007). Do children and adults with language impairment recognize prosodic cues? *Journal of Speech, Language, and Hearing Research*, *50*(3), 746–758. https://doi.org/10.1044/1092-4388(2007/052)

Frost, D., & Picavet, F. (2014). Putting prosody first — Some practical solutions to a perennial problem: The Innovalangues Project. *Research in Language*, *12*(3), 233–243. https://doi.org/10.2478/rela-2014-0002

Gilbert, A. C., Lee, J. G., Wolpert, M., & Baum, S. R. (2023). Phrase parsing in a second language as indexed by the closure positive shift: The impact of language experience and acoustic cue salience. *European Journal of Neuroscience*, *58*(8), 3838–3858. https://doi.org/10.1111/ejn.16134

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

Gordon, M., & Roettger, T. (2017). Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard*, *3*(1), 20170007. https://doi.org/10.1515/lingvan-2017-0007

Gussenhoven, C. (2008). Types of focus in English. In C. Lee, M. Gordon, & D. Buring (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 83–100). Springer.

Idemaru, K., & Holt, L. (2013). The developmental trajectory of children's perception and production of English /r/-/l/. *The Journal of the Acoustical Society of America*, *133*, 4232–4246. https://doi.org/10.1121/1.4802905

Ingvalson, E. M., Ettlinger, M., & Wong, P. C. M. (2014). Bilingual speech perception and learning: A review of recent trends. *International Journal of Bilingualism*, *18*(1), 35–47. https://doi.org/10.1177/1367006912456586

Ionin, T., Luchkina, T., & Goldshtein, M. (2024). Contrastive focus is acquirable: An investigation of Russian contrastive focus with English/Russian bilinguals. *Second Language Research*, *40*(3), 709–737. https://doi.org/10.1177/02676583231178101

Ip, M. H. K., & Cutler, A. (2016). Cross-language data on five types of prosodic focus. *Speech Prosody 2016 Proceedings*, 330–334. https://doi.org/10.21437/SPEECHPROSODY.2016-68

Jasmin, K., Dick, F., & Tierney, A. T. (2020). The Multidimensional Battery of Prosody Perception (MBOPP). *Wellcome Open Research*, *5*(4). https://doi.org/10.12688/wellcomeopenres.15607.1

Jasmin, K., Sun, H., & Tierney, A. T. (2021). Effects of language experience on domain-general perceptual strategies. *Cognition*, *206*, 104481. https://doi.org/10.1016/j.cognition.2020.104481

Jasmin, K., Tierney, A. T., Obasih, C., & Holt, L. L. (2023). Short-term perceptual reweighting in suprasegmental categorization. *Psychonomic Bulletin Review*, *30*, 373–382. https://doi.org/10.3758/s13423-022-02146-5

Jongman, A., & Tremblay, A. (2020). Word prosody in second language acquisition. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 594–604). Oxford University Press.

Kachlicka, M., Symons, A. E., Ruan, Y., Saito, K., Dick, F., & Tierney, A. T. (2025). Effects of targeted perceptual training on L2 suprasegmental weighting strategies training study. https://doi.org/10.31234/osf.io/y4uph

Kang, O., Thomson, R. I., & Moran, M. (2020). Which features of accent affectunderstanding? Exploring the intelligibility threshold of diverse accentvarieties. *Applied Linguistics*, *41*(4), 453–480. https://doi.org/10.1093/applin/amy053

Kalathottukaren, R. T., Purdy, S. C., & Ballard, E. (2015). Behavioral measures to evaluate prosodic skills: A review of assessment tools for children and adults. *Contemporary Issues in Communication Science and Disorders*, *42*, 138–154. https://doi.org/10.1044/cicsd_42_S_138

Kawahara, H., & Irino, T. (2005). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In P. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 167–180). Springer. https://doi.org/10.1007/0-387-22794-6_11

Kelly, M. H. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition*, *30*(2), 107–137. https://doi.org/10.1016/0010-0277(88)90037-6

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153–194.

Kiss, K. E. (1998). Identificational focus versus information focus. *Language*, *74*(2), 245–273. https://doi.org/10.2307/417867

Knill, K., Nicholls, D., Gales, M. J. F., Qian, M., & Stroinski, P. (2024). Speak & Improve Corpus 2025: An English speech corpus for language assessment and feedback. https://doi.org/10.48550/arXiv.2412.11986

Kolinsky, R., Cuvelier, H., Goetry, V., Peretz, I., & Morais, J. (2009). Music training facilitates lexical stress processing. *Music Perception*, *26*(3), 235–246. https://doi.org/10.1525/mp.2009.26.3.235

Kostromitina, M., & Plonsky, L. (2022). Elicited imitation task as a measure of L2 proficiency. A meta-analysis. *Studies in Second Language Acquisition*, *44*(3), 886–911. https://doi.org/10.1017/S0272263121000395

Krause, S. (1982). Vowel duration as a perceptual cue to postvocalic consonant voicing in young children and adults. *The Journal of the Acoustical Society of America*, *71*, 990–995. https://doi.org/10.1121/1.387580

Kuang, J., Chan, M. P. Y., & Rhee, N. (2022). The effects of syntactic and acoustic cues on the perception of prosodic boundaries. *Speech Prosody 2022 Proceedings*, 699–703. https://doi.org/10.21437/SpeechProsody.2022-142

Kusz, E. (2022). Effects of imitation and self-imitation practice on L2 pronunciation progress. *Topics in Linguistics*, *23*(2), 1–17. https://doi.org/10.2478/topling-2022-0008

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

Lee, G. (2022). Perceptual weighting on English lexical stress by Korean learners of English. *Phonetics and Speech Sciences*, *14*(4), 19–24. https://doi.org/10.13064/KSSS.2022.14.4.019

Lee, G., Shin, D. J., & Martinez Garcia, M. T. (2019). Perception of lexical stress and sentence focus by Korean-speaking and Spanish-speaking L2 learners of English. *Language Sciences*, *72*, 36–49. https://doi.org/10.1016/j.langsci.2019.01.002

Lengeris, A. (2012). Prosody and second language teaching: Lessons from L2 speech perception and production research. In J. Romero-Trillo (Ed.), *Pragmatics and Prosody in English Language Teaching*, 25–40. Springer Netherlands. https://doi.org/10.1007/978-94-007-3883-6_3

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signalling /b/ versus /p/ in trochees. *Language and Speech*, *29*(1), 3–11.

Liu, D. (2021). Prosody transfer failure despite cross-language similarities. Evidence in favor of a complex dynamic system approach in pronunciation teaching. *Journal of Second Language Pronunciation*, *7*(1), 38–61. https://doi.org/10.1075/jslp.18047.liu

Mayo, C., Scobbie, J. M., Hewlett, N., & Waters, D. (2003). The influence of phonemic awareness development on acoustic cue weighting strategies in children's speech perception. *Journal of Speech, Language, and Hearing Research*, *46*(5), 1184–1196. https://doi.org/10.1044/1092-4388(2003/092)

Mayo, C., & Turk, A. (2005). The influence of spectral distinctiveness on acoustic cue weighting in children's and adults' speech perception. *The Journal of the Acoustical Society of America*, *118*, 1730–1741. https://doi.org/10.1121/1.1979451

Mora-Plaza, I., Saito, K., Suzukida, Y., Dewaele, J.-M., & Tierney, A. (2022). Tools for second language speech research and teaching. http://sla-speech-tools.com. http://doi.org/10.17616/R31NJNAX

Mundy, I. R., & Carroll, J. M. (2012). Speech prosody and developmental dyslexia: reduced phonological awareness in the context of intact phonological representations. *Journal of Cognitive Psychology*, *24*(5), 560–581. https://doi.org/10.1080/20445911.2012.662341

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Ng, M. L., & Chen, Y. (2011). Proficiency in English sentence stress production by Cantonese speakers who speak English as a second language (ESL). *International Journal of Speech-Language Pathology*, *13*(6), 526–535. https://doi.org/10.3109/17549507.2011.580783

Nittrouer, S. (2002). Learning to perceive speech: how fricative perception changes, and how it stays the same. *The Journal of the Acoustical Society of America*, *112*, 711–719. https://doi.org/10.1121/1.1496082

Nittrouer, S., & Lowenstein, J. H. (2015). Weighting of acoustic cues to a manner distinction by children with and without hearing loss. *Journal of Speech, Language, and Hearing Research*, *58*(3), 1077–1091. https://doi.org/10.1044/2015_JSLHR-H-14-0263

O'Brien, M. G., Jackson, C. N., & Gardner, C. E. (2014). Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. *Applied Psycholinguistics*, *35*(1), 27–70. https://doi.org/10.1017/S0142716412000252

Ordin, M., Polyanskaya, L., Laka, I., & Nespor, M. (2017). Cross-linguistic differences in the use of durational cues for the segmentation of a novel language. *Memory & Cognition*, *45*(5), 863–876. https://doi.org/10.3758/s13421-017-0700-9

Peng, S. C., Lu, N., & Chatterjee, M. (2009). Effects of cooperating and conflicting cues on speech intonation recognition by cochlear implant users and normal hearing listeners. *Audiology and Neurotology*, *14*(5), 327–337. https://doi.org/10.1159/000212112

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress "deafness". In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7*. De Gruyter Mouton.

Peperkamp, S., Vendelin, I., & Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, *38*(3), 422–430. https://doi.org/10.1016/j.wocn.2010.04.001

Peppé, S., Maxim, J., & Wells, B. (2000). Prosodic variation in Southern British English. *Language and Speech*, *43*(3), 309–334. https://doi.org/10.1177/00238309000430030501

Petrova, K., Jasmin, K., Saito, K., & Tierney, A. T. (2023). Extensive residence in a second language environment modifies perceptual strategies for suprasegmental categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *49*(12), 1943–1955. https://doi.org/10.1037/xlm0001246

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

Pinter, G., Mizuguchi, S., & Tateishi, K. (2014). Perception of prosodic prominence and boundaries by L1 and L2 speakers of English. *Conference of the International Speech Communication Association (INTERSPEECH)*, 544–554. https://www.isca-archive.org/interspeech_2014/pinter14_interspeech.pdf

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech*, *54*(2), 147–165. https://doi.org/10.1177/0023830910397489

Repp, B. H. (1983). Coarticulation in sequences of two nonhomorganic stop consonants: Perceptual and acoustic evidence. *The Journal of the Acoustical Society of America*, *74*(2), 420–427. https://doi.org/10.1121/1.389835

Rojczyk, A. M., & Porzuczek, A. (2012). Selected aspects in the acquisition of English phonology by Polish learners – segments and prosody. In D. Gabryś-Barker (Ed.), *Readings in second language acquisition* (pp. 93–120). Wydawnictwo Uniwersytetu Śląskiego.

Rooth, M. E. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*, 75–116. https://doi.org/10.1007/BF02342617

Shin, J., & Speer, S.R. (2012). English lexical stress and spoken word recognition in Korean learners of English. *6th International Conference Speech Prosody*, 422–425. https://doi.org/10.21437/SpeechProsody.2012-107.

Sun, H., Saito, K., & Tierney A. T. (2021). A longitudinal investigation of explicit and implicit auditory processing in L2 segmental and suprasegmental acquisition. *Studies in Second Language Acquisition*, *43*(3), 551–573. https://doi.org/10.1017/S0272263120000649

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning*, *65*(4), 860–895. https://doi.org/10.1111/lang.12138

Tremblay, A., Broersma, M., Zeng, Y., Kim, H., Lee, J., & Shin, S. (2021). Dutch listeners' perception of English lexical stress: A cue-weighting approach. *The Journal of the Acoustical Society of America*, *149*, 3703–3714. https://doi.org/10.1121/10.0005086

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*(1), 1–30. https://doi.org/10.1017/S0272263106060013

Trouvain, J., & Braun, B. (2020). Sentence prosody in a second language. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 605–618). Oxford University Press.

Tseng, C., & Su, C. (2014). Where and how to make an emphasis? L2 distinct prosody and why. *The 9th International Symposium on Chinese Spoken Language Processing*, 633–637. https://doi.org/10.1109/ISCSLP.2014.6936585

Vander Klok, J., Goad, H., & Wagner, M. (2018). Prosodic focus in English vs. French: A scope account. *Glossa: A Journal of General Linguistics*, *3*(1), 71. https://doi.org/10.5334/gjgl.172

Peter, V., McArthur, G., & Crain, S. (2014). Using event-related potentials to measure phrase boundary perception in English. *BMC Neuroscience*, *15*(129). https://doi.org/10.1186/s12868-014-0129-z

Wang, X. (2020). Segmental versus suprasegmental: Which one is more important to teach? *RELC Journal*, *53*(1), 194–202. https://doi.org/10.1177/0033688220925926

Wierzchowska, B. (1971). *Wymowa Polska*. Panstwowe Wydawnictwo Naukowe.

Wu, S. L., Tio, Y. P., & Ortega, L. (2021). Elicited imitation as a measure of L2 proficiency. *Studies in Second Language Acquisition*, *44*(1), 271–300. https://doi.org/10.1017/S0272263121000103

Yu, J., & Gibbon, D. (2015). How natural is Chinese L2 English prosody? *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0304.pdf

Yu, V. Y., & Andruski, J. E. (2009). A cross-language study of perception of lexical stress in English. *Journal of Psycholinguistic Research*, *39*, 323–344. https://doi.org/10.1007/s10936-009-9142-2

Zerbian, S. (2015). Markedness considerations in L2 prosodic focus and givenness marking. In: E. Delais-Roussarie, M. Avanzi, & S. Herment (Eds.), *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations*. Springer-Verlag.

Zhang, X. (2012) *A comparison of cue-weighting in the perception of prosodic phrase boundaries in English and Chinese.* PhD thesis, The University of Michigan.

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press

Zhang, Y., Ding, H., Zelchenko, P., Cui, X., Lin, Y., Zhan, Y., & Zhang, H. (2018). Prosodic disambiguation by Chinese EFL learners in a cooperative game task. *9th International Conference on Speech Prosody 2018 (Poznan, Poland)*, 979–983. https://doi.org/10.21437/SpeechProsody.2018-198

Zhang, Y., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America*, *123*(6), 4498–4513. https://doi.org/10.1121/1.2902165

Cite this article: Kachlicka, M., Jasmin, K., Symons, A., Saito, K., Dick, F., & Tierney, A. (2026). English Prosody Dataset for investigating English prosody acquisition. *Studies in Second Language Acquisition*, 1–20. https://doi.org/10.1017/S0272263126101788

https://doi.org/10.1017/S0272263126101788 Published online by Cambridge University Press