The Musical Impact of Multicultural London English (MLE) Speech Rhythm

Christopher S. Lee, Lucinda Brown, &
Daniel Müllensiefen
University of London, London, United Kingdom

There is evidence that an emerging variety of English spoken by young Londoners—Multicultural London English (MLE)—has a more even syllable rhythm than Southern British English (SBE). Given findings that native language rhythm influences the production of musical rhythms and text setting, we investigated possible musical consequences of this development. We hypothesized that the lower vocalic durational variability in MLE and (putatively) less salient stress distinctions would go along with a preference by MLE speakers for lower melodic durational variability and a higher tolerance for stress mismatches (the non-coincidence of stress/beat strong-weak patterns) compared to SBE speakers. An analysis of two popchart song corpora by MLE and SBE artists confirmed that durational variability was lower in the MLE songs, and that there were more stress mismatches. In a follow-up experiment, MLE and SBE participants read four short English sentences and then rated text settings in pairs of specially constructed song fragments with and without stress mismatches. MLE participants' speech showed the expected lower variability in vocalic duration and syllabic prominence compared to SBE participants' speech, while their text setting ratings showed a greater tolerance of stress mismatches.

T here is now a substantial body of evidence that the rhythmic properties of a language are reflected in music produced by native speakers of the language, both in the durational patterning of melodic rhythms, and in the way in which the syllables of a text are aligned with the beats of the musical metre when a text is set to music. A pioneering study by Patel and Daniele (2003a; see also Patel, Iversen, & Rosenberg, 2006), and subsequent studies by Huron and Ollen (2003), McGowan and Levitt (2011), and Sadakata (2006) yielded the finding that the degree of vocalic durational variability in a language is reflected in the degree of durational variability in the melodies produced by native speakers, across a wide variety of dialects, languages, and musical styles, though more detailed findings from European art music have also shown how stylistic factors (in this case probably reflecting non-native language rhythm) may exert an overriding counterinfluence (Daniele & Patel, 2004, 2013; Hansen, Sadakata, & Pearce, 2016; Huron & Ollen, 2003; Patel & Daniele, 2003b; VanHandel & Song, 2010). These similarities appear to be salient to listeners: Hannon (2009) showed that native and non-native English listeners were able to classify French and English folk songs according to their language of origin solely on the basis of their rhythms. With regard to text setting, Dell and Halle (2009) observed that stress mismatches, in which a strong-weak/weak-strong syllabic stress pattern is aligned with the opposite metrical pattern, are commonly found in French songs but not in English songs, and Temperley and Temperley (2013) provided quantitative confirmation of their observation in a study in which they compared text settings in two corpora of traditional English and French songs. Dell and Halle suggested that tolerance of stress mismatches depends on the perceptual salience of linguistic stress in the language concerned, as determined by phonological factors such as the presence or absence of vowel reduction: they claim that mismatches are less tolerated in English than in French because linguistic stress is far more salient in English (see also Rodriguez-Vazquez, 2010, for similar observations and claims regarding English as compared to Spanish text setting). Their claim is supported by a recent auditory model-based analysis of two multilingual speech corpora (Lee & Todd, 2004; see also Lee, Kitamura, Burnham, & Todd, 2014), which showed that English syllables vary more in prominence than French syllables. The auditory-model account yields the further more general claim that prevalence of stress mismatches and degree of melodic durational variability may go hand in hand, because the underlying prosodic phenomena are linked: vowel length is an important determinant of syllabic prominence, and

TABLE 1. *Songs Used in the Analysis, by Artist Group (MLE or SBE), Artist, and Title*

MLE artists		SBE artists	
Artist (year of birth)	Song (year, peak chart position)	Artist (year of birth)	Song (year, peak chart position)
Dappy (1987)	No regrets (2011, 1)	Sophie Ellis-Bextor (1979)	Move this mountain (2002, 3)
			Murder on the dance floor (2002, 2)
Jessie J (1988)	Do it like a dude (2010, 2)	Tom Fletcher (1985)	Five colours in her hair (2004, 1)
	Domino (2011, 1)		
	Laserlight (2012, 5)		
	Price tag (2011, 1)		
Katy B (1989)	Crying for no reason (2014, 5)	Ellie Goulding (1986)	Anything could happen (2012, 5)
	Katy on a mission (2010, 5)		Starry eyed (2010, 4)
			The writer (2010, 19)
Labrinth (1989)	Beneath your beautiful (2012, 1)	Ed Sheeran (1991)	Drunk (2012, 9)
	Earthquake (2011, 2)		Lego house (2011, 5)
	Let the sun shine (2010, 3)		The A team (2011, 3)
			You need me, I don't need you (2011, 4)
Ms Dynamite (1981)	Dy-na-mi-tee (2002, 5)	Florence Welch (1986)	Rabbit heart (2008, 12)
	Put him out (2002, 19)		Shake it out (2011, 12)
			Spectrum (2012, 1)
Plan B (1983)	Prayin' (2010, 16)	Will Young (1979)	Don't let me down (2002, 2)
	She said (2010, 3)		Jealousy (2011, 5)
	Stay too long (2010, 9)		Your game (2004, 3)
	The recluse (2010, 35)		

hence a high degree of vocalic durational variability will tend to lead to more salient stress contrasts.[1]

In the study presented here, we investigate the musical impact of the recent emergence of a rhythmically distinct variety of English, Multicultural London English (MLE). MLE is spoken by young inner-city Londoners of different ethnicities, and emerged from the contact of various ethnic and second-language varieties of English (including Caribbean English) in the 1980s and 1990s (Cheshire, Kerswill, Fox, &Torgersen, 2011). It is rhythmically different from standard varieties of English, such as Southern British English (SBE), in that it has a more even syllable rhythm (Torgersen & Szakay, 2012). We hypothesize that songs written by MLE speakers contain less melodic durational variability and more stress mismatches than songs written by SBE speakers, and that this reflects MLE speakers' greater tolerance of stress mismatches and the lower degree of vocalic durational variability and variability in syllabic prominence in their speech compared to SBE speakers. In order to test these hypotheses, we first analyze two song corpora of popchart songs, one co-written

and performed by MLE artists and the other co-written and performed by SBE artists, and compare their degree of melodic durational variability and the prevalence of stress mismatches. We then investigate the text setting intuitions of MLE and SBE speakers and the degree of vocalic durational variability and variability in syllabic prominence in their speech in an experiment in which they read 4 short English sentences and then rate the goodness of fit between text and music in pairs of specially constructed song fragments with or without stress mismatches.

Experiment 1: Song-Corpora Analysis

MATERIALS

Thirty-two popchart songs (see Table 1) were selected for analysis, 16 co-written by SBE artists (3 males, 3 females), and 16 by MLE artists (3 males, 3 females). While it would have been preferable to have songs that were not co-written in order to avoid diluting any effect of the chosen artist's native dialect rhythm, this proved impossible given the trend in recent years towards multiple collaborations (see, e.g., Kopf, 2015), and all but three songs (Labrinth's "Let the sun shine," and Ed Sheeran's "The A team" and "You need me, I don't need you") had multiple songwriting credits. All songs (except Katy B's "Crying for no reason") reached the top 100 in the UK singles chart between 2001 and 2012,

---

[1] The relationship between vowel length and prominence breaks down, however, in the case of linguistic phenomena such as the bisyllabic words with an initial stressed but very short syllable that occur in English (and less commonly, German). These are reflected in the "Scotch snap" rhythmic figure often found in British songs in the 17th, 18th, and 19th centuries (Temperley & Temperley, 2011).

and were the most popular (measured by peak chart position) during that period of those songs by the artist concerned for which sheet music scores were available.[2] Artists were selected on the basis of the following criteria: 1) their accent (see Appendix A), in interviews recorded in 2012; 2) the location where they grew up (South-East England, including London, for the SBE artists, and London for the MLE artists); 3) their commercial success (all of them had co-written at least two songs reaching the top 100 in the UK singles chart between 2001 and 2012); 4) their age (all of them were under 35 in 2012); and 5) their monolingualism (none of them, as far as we could determine, were fluent speakers of any language other than English). Sheet music scores for each song were obtained via www.sheetmusicplus.com, www.musicnotes.com, and various printed compilations.

METHOD

Analyses were performed using the vocal lines of whole songs, taken from the sheet music scores. Certain song passages/sections were excluded, in order to avoid a number of potential confounding factors: 1) sections without a melodic line (rap sections), or largely devoid of text (intro/outro sections containing mostly wordless melismas or repeated riffs), because of their potentially different rhythmic characteristics and varying distribution across the two corpora; and 2) repeats of passages of two bars or more (passages with the same melodic rhythm—disregarding ornamentation—and text as an earlier passage), because of possible variability in the amount of repetition and in the rhythmic characteristics of repeated material across the two corpora. We also excluded sections in which the underlay of the lyrics was not entirely free of rhythmic ambiguities (e.g., second and subsequent verses without a fully written-out melodic line), as well as passages different/absent from the official YouTube recorded versions. Note durations exceeding a bar were trimmed to a bar's length, while in two MLE songs, rhythms indicated as having a swing/shuffle feel but not notated as such were reanalyzed in a 2:1 ratio.[3]

Two analyses were carried out, comparing: 1) the degree of melodic durational variability across the two corpora, using the normalized Pairwise Variability Index (nPVI; Grabe & Low, 2002; Patel & Daniele, 2003a); and 2) the prevalence of stress mismatches across the two corpora.

nPVI analysis. With the exception of VanHandel and Song (2010; see also VanHandel, 2005) and Sadakata (2006; see also Sadakata, Desain, Honing, Patel, & Iversen, 2004), previous nPVI analyses have used single themes or short phrases typically not exceeding four measures in length. As noted above, however, the analysis here took the vocal lines of whole songs, since in the absence of a dictionary of themes or equivalent resource, there is no nonarbitrary way of identifying representative segments for analysis. One important issue that then arises is how to deal with rests: most particularly, how to distinguish within-phrase rests from between-phrase rests. A within-phrase rest functions as an articulation marker and hence should arguably be added to the duration of the preceding note and/or not prevent the preceding and following notes from forming a note pair (Daniele & Patel, 2013; Huron & Ollen, 2003; VanHandel & Song, 2010), while a between-phrase rest marks a boundary between phrases, and hence the preceding and following notes do not form a note pair and their durations remain unchanged. The most thoroughgoing way to distinguish between the two types is on the basis of a prior phrasal segmentation analysis (the approach taken by VanHandel and Song), but given the difficulties of such an approach, we opted for a simpler solution: we treated all rests as between-phrase rests and excluded short segments (continuous runs of fewer than seven notes, following Daniele and Patel, 2004) from the analysis.[4] The rationale is as follows: 1) short segments are more likely to be followed by within-phrase rests and hence to yield misleading nPVI values than long segments, so their exclusion is likely to result in more valid nPVI estimates;[5] 2) there may be differences between songs and song corpora in the prevalence of short and long segments, so the elimination of short segments is also likely to yield more valid comparative nPVI estimates.

---

[2] "Crying for no reason" reached the top 100 in 2014, and was included due to the unavailability of sheet music scores for other songs by Katy B apart from "Katy on a mission," and for any extra songs by fellow MLE artists, Dappy, Labrinth, and Ms Dynamite.

[3] The rhythms reanalyzed were the even eighth notes in sections of Plan B's "She said," and the sequences of dotted sixteenth and thirty-second notes in Ms Dynamite's "Dy-na-mi-tee."

[4] Sadakata and colleagues calculate nPVI values from the note interonset intervals (IOIs) in MIDI files, rather than from the musical notation in sheet music scores, and so the notational issue does not arise. They treat IOIs longer than a bar as phrase boundary markers and exclude them from the analyses (see Sadakata, 2006, for details).

[5] Very long segments are likely to contain one or more phrase boundaries unmarked by rests, yielding spurious note pairs straddling the boundaries, and therefore also result in misleading nPVI values. However, it could be argued that each such case will only yield one error, whereas a within-phrase rest yields two errors: a missing nPVI value for the note pair either side of the rest, and an incorrect note value for the segment-final note preceding the rest.

Hence, the nPVI of each song was simply calculated as the mean of the nPVIs of every note pair in a continuous sequence of at least seven notes not separated by a rest (ignoring grace notes and fermatas).

Stress mismatch analysis. The analysis involved 1) selecting a suitable set of pairs of stressed and unstressed syllables; 2) determining the metrical weight (Longuet-Higgins & Lee, 1984) of the notes on which each syllable occurred, taking into account the effects of metrical displacements and polyrhythms (see below). Melismatic syllables (associated with more than one note), and hence any syllable pair containing melismatic syllables, were excluded from the analysis. A stress mismatch was defined as the occurrence of a syllable pair in which the stressed syllable falls on a note of lesser weight than the unstressed syllable, and a stress match as the occurrence of a pair in which the stressed syllable is at least as heavy as the unstressed syllable. Matches and mismatches were classified as to whether the tactus-level was involved (i.e., where one event in a syllable pair occurs on a tactus-level beat, but the other does not), as it has been claimed (e.g., Dell & Halle, 2009) that tactus-level mismatches are more unacceptable than those involving other metrical levels.

Choice of syllable pairs. From the set of all syllable pairs that would yield stress mismatches if paired with a contradictory metrical stress pattern, a subset of pairs was selected, consisting of consecutive syllables of clearly different stress status in tightly bound prosodic domains, which evidence suggests would yield particularly unacceptable mismatches unlikely to occur in standard English text settings (see Dell & Halle, 2009; Hayes, 2009; Hayes & Kaun, 1996, on the influence of prosodic grouping on the acceptability of stress mismatches). Pairs met the following criteria: 1) one syllable is unstressed, and the other bears primary or secondary stress; 2) both syllables are in the same word, or the unstressed syllable is a highly unstressable function word (article "a(n)/the," preposition/infinitive particle "for, of, to") preceding a stressed syllable in a following monosyllabic content word (adjective, noun, or lexical verb) or multisyllabic word (see Littlefield, 2006, on prepositions, and Shih, 2014, on stressability and the function/content word distinction).

We checked the stress pattern of polysyllabic words for both standard Southern British English using the Longman Pronunciation Dictionary (Wells, 2008), and also General American (using the online Carnegie Mellon Pronouncing Dictionary; Carnegie Mellon University, 2008); any syllable pairs thereby found to have more than one possible stress pattern (within or across the two dialects) were excluded. This likely yields a highly reliable check on the SBE artists' stress patterns, though it is not clear to what extent it yields a similarly reliable check on the MLE artists' stress patterns. However, given evidence from a comparison of SBE and General American (Berg, 1999) that dialectal variation in lexical stress patterns (at least in English) may be fairly limited (around 1.2% of the 75,000 word corpus in Berg's study), and predominantly confined to longer, less frequent words, it is likely that the same procedure is also valid for the MLE artists, especially since: 1) only 64 out of the total of 472 syllable pairs (13.6%) included in the analysis of the MLE songs were part of a word (lemma)[6] of more than two syllables; and 2) only 34 syllable pairs (7.2%) were part of a word outside the top 5,000 most frequent words in the Corpus of Contemporary American English (Davies, 2008).

Metrical analysis. As Temperley (1999, 2001) has noted, popular Western styles of music are full of various types of metrical displacement, which involve surface deviations from an underlying metrical representation in which the deviation is resolved, and it is the stress pattern of the underlying representation that determines whether a text setting is heard as mismatched. In addition, as noted by Liberman (2007, 2009), there are polyrhythms, where there is no metrical displacement but a temporary change in stress pattern as a result of the occurrence of a rhythmic figure implying a different metre, and these too seem to affect mismatch judgments. On the basis of their accounts, we developed a procedure for metrical analysis that allowed for the displacements and polyrhythmic reinterpretations they describe, guided by the general assumption that the preferred analysis is one that yields stress matches rather than mismatches. The procedure is described in detail in Appendix B.

RESULTS

The mean nPVI of the vocal lines of the SBE songs was 40.5 ($SD = 9.0$), compared to 31.2 ($SD = 14.1$) for the MLE songs (see Table A, Appendix B, for detailed results); the difference yielded a significant effect on a two-tailed unrelated samples t-test, $t(30) = 2.20$, $p = .04$; Cohen's $d = 0.79$.

The results of the stress mismatch analysis are shown in Table 2. The proportion of stress mismatches was

---

[6] All word frequency estimates are based on lemmas (canonical forms), which group together morphological variants of the same word (so that e.g., "voices" is an instance of the monosyllabic lemma "voice").

TABLE 2. *Total Number of Candidate Pairs, Number of Stress Matches and Mismatches at all Metrical Levels/Tactus-level Only in the MLE and SBE Song Corpora*

Song corpus	Total candidate pairs	Stress matches	Stress mismatches
MLE	472/376	457/364 (96.8%/96.8%)	15/12 (3.2%/3.2%)
SBE	464/314	458/312 (98.7%/99.4%)	6/2 (1.3%/0.6%)

marginally significantly higher overall in the MLE than the SBE songs (3.2% vs. 1.3%; $p = .08$, Fisher's exact p value), and significantly higher in the case of tactus-level stress mismatches (3.2% vs. 0.6%; $p = .03$, Fisher's exact p value).

DISCUSSION

The results of the analyses confirm that the MLE songs have less melodic durational variability and contain more stress mismatches than the SBE songs, in accordance with our hypotheses. The durational variability findings suggest a robust effect of native dialect rhythm, despite the multiple authorship of most of the songs, though the stress mismatch findings are much weaker, with only tactus-level mismatches showing a statistically reliable effect. However, the durational variability findings may be compromised by the confound of a possible stylistic influence of rap, since the fact that rap sections of songs were excluded from analysis does not control for any more general stylistic effect. It seems likely that rap has less durational variability than other vocal styles (though we are not aware of any published findings), and if its stylistic influence extends to melodic rhythms more generally, then the effect is likely to be greater on MLE songs/artists than SBE songs/artists: 1) only one of the SBE songs (Ed Sheeran's "You need me, I don't need you") contains a rap section, as against 6 of the MLE songs (Dappy's "No regrets," Jessie J's "Price tag," "Labrinth's "Earthquake," Plan B's "She said," "Stay too long," and "The recluse"); 2) all the MLE artists are either rappers (Dappy, Ms Dynamite, Labrinth, and Plan B) or have collaborated with rappers (Jessie J with US rapper B.o.B., and Katy B with Ms Dynamite), while of the SBE artists, only Ed Sheeran is a rapper, and only Ellie Goulding and Florence Welch have collaborated with rappers (Ellie Goulding with English rapper Tinie Tempah, and Florence Welch with US rapper ASAP Rocky, and English rapper Dizzee Rascal).

We therefore reanalyzed the nPVI results to try and determine the putative confounding influence of rap, either at the level of the individual song, or the artist. With regard to individual songs, there is some evidence that songs with rap sections have a lower nPVI than songs without rap sections: the one SBE song with a rap section (Ed Sheeran's "You need me, I don't need you") has a much lower nPVI (27.1) than the group average, while the MLE songs with a rap section have a slightly lower mean nPVI (28.4; $SD = 13.2$) than the group average. However, removing these songs has little effect on the overall means of the two groups, only slightly raising the mean nPVI of both the SBE songs (41.3; $SD = 8.6$) and MLE songs (33.0; $SD = 15.0$), and reducing the between-group effect size as measured by Cohen's d from 0.79 to 0.69. With regard to individual artists, there is also evidence of a rap influence: the songs of the rap-influenced SBE artists (Ellie Goulding, Ed Sheeran, Florence Welch) have a lower mean nPVI (38.3; $SD = 9.9$) than those of the other SBE artists (Sophie Ellis-Bextor, Tom Fletcher, Will Young; mean nPVI = 44.1; $SD = 6.4$). However, their mean nPVI is still well above the mean for the MLE songs (38.3 vs. 31.2), though the removal of the songs by the other SBE artists results in a reduction of the between-group effect size as measured by Cohen's d to 0.59. In short, there is suggestive evidence of a stylistic effect, but not so large as to call into question the main finding. However, given the small size of our song corpora, the relative importance of linguistic and stylistic influences remains to be more precisely determined.

With regard to the stress mismatch findings, the difference between the two groups, as already noted, is small. The rate of stress mismatching in the MLE songs, at 3.2% is quite low, particularly when compared to the rates of around 35-40% reported by Temperley and Temperley (2013) in their corpus of traditional French songs. Although caution is required in comparing the MLE and French rates given differences in the methods of analysis and stylistic characteristics of the song corpora, the much lower rate in our MLE corpus could reflect the likely continuing importance of lexical stress for spoken word recognition in MLE, as in other varieties of English, compared to a fixed-stress language like French where it plays no role (see Cutler, 2005, for a review). It also could suggest that the perceptual salience of stress is not the only factor determining the acceptability of stress mismatching. However, there are a number of other possible explanations, including the possibility that it could be an artefact of our method of

analysis, and in particular of the assumption that listeners choose stress-matched interpretations wherever possible. We consider the issue in more detail in the General Discussion.

In short, the findings relating to melodic durational variability seem fairly solid, despite the likely confounding influence of stylistic factors, but the stress mismatch findings are weaker, possibly due to a variety of additional factors that might be obscuring the stress-matching preferences of the MLE songwriters. In the next section, we describe a follow-up experiment, which uses a more direct method to determine MLE and SBE speakers' stress-matching preferences: it directly probes their text setting intuitions, using a novel paradigm in which they rate text settings in pairs of specially constructed song fragments with and without stress mismatches. The experiment also investigates the variability in syllabic prominence (as well as vocalic duration) in their speech, in order to test the hypothesis that tolerance of stress mismatches reflects less salient stress distinctions.

Experiment 2: Speech Rhythm and Text Setting Rating Experiment

PARTICIPANTS

Nineteen SBE speakers (11 males, 8 females) and 14 MLE speakers (7 males, 7 females) took part in the experiment. Participant age ranged from 18-30 years, and all were monolingual (defined as answering no to the question "do you speak any language other than English fluently?"). The SBE speakers had been raised and were currently residing in South-East England (including London), while the MLE speakers had been raised and were currently residing in London. Both groups were selected by opportunity sampling (via personal contacts in the case of the SBE speakers, and via interception at the Westfield Stratford City shopping centre, Stratford, East London, in the case of the MLE speakers). Their accents were initially identified via casual conversation and later confirmed after listening to their readings of four test sentences (but prior to the formal rhythmic analysis of the readings).

TEST SENTENCES AND ANALYSIS OF PARTICIPANTS' SPEECH RHYTHM

Four short test sentences (see Appendix A) were recorded by participants on a laptop using Windows Sound Recorder; participants were instructed to read the sentences at a normal conversational speed, without pauses or hesitations. The sentences were designed to elicit key segmental features of the two accents as an aid to accent identification, and also to have roughly equal proportions of stressed and unstressed syllables so as to provide a good test of hypothesized rhythmic differences.

After elimination of a small number of readings that were unusable due to errors or hesitations, one sentence remained with satisfactory readings by all participants ("they cost a lot of money, so no-one buys them") and was therefore selected for analysis. The connective "so" was ignored in all analyses, due to large interspeaker differences in realization (from single consonant to separate intonational phrase).[7]

The sentences were low-pass filtered and rescaled to a standard root mean squared amplitude (RMS) value, and then segmented into consonantal and vocalic intervals using *Praat* (Boersma & Weenink, 2009). From this analysis the standard rhythm metrics were derived: the ratio of vocalic to consonantal material, $\%V$, the global measures of vocalic and consonantal durational variability, ΔV, ΔC, VarcoV, and VarcoC, and the measures of serial variability, nPVI-V and rPVI-C (Grabe & Low, 2002; Ramus, 2002; Wiget, White, Schuppler, Grenon, Rauch, & Mattys, 2010), as well as the syllable rate (syllables per second). We also calculated the peak loudness level of the syllabic nuclei, using the *Praat* loudness model, together with their mean RMS intensity, and derived measures of global and serial variability (the standard deviations of the mean intensities and maximum loudness levels across an utterance, ΔI and ΔL_{max}, and the raw PVI equivalents, rPVI-I and rPVI-L_{max}), following the method in Lee et al. (2014). Further details of the analysis are given in Appendix C.

TEXT SETTING RATING TASKS: MATERIALS

Twelve song fragments were composed for the rating tasks, using Sibelius software (2 for the practice trials, and the remaining 10 for the experimental trials). Each contained a 4-bar instrumental introduction (preceded by a 1-bar click track), a 2-bar vocal phrase, and a final 2-bar instrumental break; half were in pop ballad style (90 BPM), and half in eighth-note feel rock style (120 BPM), with piano, bass, and drums accompaniment. The vocals were sung by a London-based professional jazz/contemporary vocalist. There were two versions of each fragment: 1) in the original stress-matched versions, consecutive stressed and unstressed syllables occurred on consecutive eighth-note beats on the tactus and subtactus respectively; 2) in the stress-mismatched versions, the original recorded vocals of the first half of the phrase (terminated by an eighth-note rest) were shifted an eighth-note later, thereby yielding

---

[7] Data was missing for one MLE female participant.

one or more tactus-level stress mismatches (according to the same criteria as those used in the song corpora analysis). The words in the lyrics were short (no words of more than two syllables) and high frequency, thereby virtually eliminating the possibility of between-dialect variation in stress patterns confounding the results (see earlier discussion of the song corpora lyrics).

The fragments were constructed in such a way as to minimize the possibility of factors other than the surface metrical alignment of the melody influencing participants' judgements. First, the melodic lines in the first half of the phrase of every fragment were monotonic and isochronous, to avoid the possibility of phenomenal accents due to pitch changes or durational factors facilitating one version over the other. Second, the eighth-note shift of the first half of the melodic line in the stress-mismatched versions eliminated the possibility of resolving the stress mismatches via a metrically displaced interpretation (i.e., an eighth note later), because that would now entail also shifting the second half of the phrase, thereby creating further stress mismatches. Figure 1 shows the critical portions of the scores of the two versions of one of the experimental fragments (song fragment 1), with the stress-mismatched syllables in the stress-mismatched version followed by question marks, together with the vocal line of a metrically displaced interpretation of the mismatched version, with the original stress mismatch now resolved, but at the cost of multiple new mismatches (mismatched syllables followed by question marks).

Table 3 gives the lyrics of all the experimental fragments; the complete set of scores is given in Appendix D.

TEXT SETTING RATING TASKS: PROCEDURE

There were two rating tasks, which used different experimental procedures to measure subjective preference for the text settings of the song fragments. In the first task, participants listened to the song fragments, and for each fragment responded to the question "How well do you think the words are set to the music?" using a 7-point scale (from *very bad* to *very good*). There were 4 practice trials before the 20 experimental trials. Half of the MLE participants and 10 out of 19 SBE participants were presented with one ordering of the experimental trials, while for the remaining participants the ordering of the stress-matched and stress-mismatched versions was reversed. The second task employed a 2AFC procedure, where participants listened to all 10 song fragments again, but this time the two versions of each song fragment were presented in immediate succession, separated by a short silent gap. Participants were then asked to select the version they thought "has the better word-setting." Half of the MLE participants and 10 out of 19 SBE participants were presented with one within-trial ordering of stress-matched and stress-mismatched versions, and the remaining participants with the reverse ordering.

RESULTS

Speech rhythm analysis. The results are shown in Table 4. The MLE sentences showed significantly lower variability on both loudness measures: $\Delta L_{\max}$, $t(30) = 2.76$; $p = .01$; rPVI-$L_{\max}$, $t(30) = 2.16$; $p = .04$, and also the nPVI measure, $t(30) = 2.46$; $p = .02$ (the contrary result on ΔV was due to the faster syllable rate of the SBE sentences, as the rate-normalized measure varcoV shows a nonsignificant difference in the same direction as the nPVI measure).

Text setting rating tasks. Difference scores were derived from the raw data from the first task (the differences in rating for the stress-matched and stress-mismatched version of each song fragment). In order to be able to average participants' ratings across the song fragment items, it was necessary to assess first whether all song fragments constitute a sufficiently homogenous item set measuring the single attribute of interest: degree of preference for (mis)matched text setting. To this end, rating data from all song fragments were submitted to a principal components analysis testing whether the rating data from all 10 song fragments could be summarized by a single component. The analysis revealed that this was not the case for the full 10-item set, and it was necessary to exclude 4 of the items (song fragments 3, 4, 9, and 10) to achieve adequate homogeneity (measure of sampling adequacy = 0.77, "middling adequacy," Kaiser & Rice, 1974) and reliability (Cronbach's alpha = 0.73) of the test set of items (see Appendix E for details of the pre-processing procedure). Once the homogeneity of the stimulus test set had been established, the rating data were averaged across participants and then used as input for the subsequent analyses.

On the reduced six-item set, in the first task the MLE participants had a mean difference score of only 0.45 ($SD = 0.98$; $MDN = 0.17$), while the SBE participants had a mean score of 1.23 ($SD = 1.32$; $MDN = 1.00$), with 6 of the 14 MLE participants showing zero or negative mean scores, as against only 3 of the 19 SBE participants. A Mann-Whitney U-test showed the difference between the two groups was significant ($p_{exact} = .04$). In the second task, the MLE participants chose the stress-matched version in only 67.9% of cases (57/84), compared to 82.5% (94/114) for the SBE participants; the difference is significant, $\chi^2(1) = 7.23$, $p = .007$.

FIGURE 1. Critical portions of the score of the stress-matched version of experimental song fragment 1 (upper panel) and the stress-mismatched version (middle panel; stress-mismatched syllables followed by question marks), together with a metrically displaced interpretation of the vocal line of the stress-mismatched version (lower panel), in which the original stress mismatch has been resolved, but at the cost of multiple new mismatches (stress-mismatched syllables followed by question marks).

TABLE 3. *Lyrics of the Experimental Song Fragments*

1. If there's *a REAson*, just TELL me now, just TELL me now.
2. LOVE is ALL *aROUND*, and the SUMMer's just beGUN.
3. GOOD *MORNing DARLing*, just CALLED to FIND OUT HOW you ARE.
4. NIGHT is *FALLing*, and I'm so FAR aWAY from HOME.
5. She's *FEELing* SAD, cos she's LOSing her WAY.
6. TELL me if you're *READy*, and TELL me if you're STRONG.
7. *PEOPle TELLing* ME, and there's PEOPle TELLing YOU.
8. *WAITing* on *the CORner*, I'm WATCHing ALL the PEOPle PASSing BY.
9. When there's *NOthing*, to EASE the PAIN of LOVE.
10. You're TIRED *of LIFE*, and you're ALL of SEVenTEEN.

Note: Stressed syllables are in upper case, and syllables in candidate pairs in the first phrase (before the comma) meeting the criteria used in the analysis are italicized. The alternating stressed/unstressed syllables in the second phrase yield multiple stress mismatches in metrically displaced interpretations of the stress-mismatched versions (i.e., shifted forward by an eighth note).

TABLE 4. *Means (SDs) for the Rhythm Metrics, Intensity Measures, Loudness Measures, and Syllable Rate by Participant Group*

	MLE	SBE	MLE-SBE
Rhythm Metrics			
%V	56.1 (3.5)	53.2 (4.8)	*ns*
ΔV	65.3 (15.8)	54.7 (10.2)	*sig*
ΔC	44.1 (10.9)	42.1 (11.5)	*ns*
VarcoV	0.57 (0.08)	0.61 (0.08)	*ns*
VarcoC	0.44 (0.07)	0.47 (0.09)	*ns*
nPVI-V	43.4 (11.3)	54.3 (13)	*sig*
rPVI-C	58.1 (13.5)	54.1 (22)	*ns*
Intensity Measures			
ΔI	3.06 (1.07)	3.31 (0.85)	*ns*
rPVI-*I*	3.62 (1.24)	4.1 (1.05)	*ns*
Loudness Measures			
ΔL_{max}	3.74 (1.08)	4.79 (1.05)	*sig*
rPVI-L_{max}	4.85 (1.16)	5.85 (1.37)	*sig*
Syllable Rate	5.5 (0.95)	6.94 (0.97)	*sig*

Note: Statistical significance for MLE versus SBE is shown in the final column.

DISCUSSION

The results of the experiment confirm our hypotheses regarding both participants' speech rhythm and their text setting preferences. MLE participants displayed lower variability in syllabic prominence as well as vocalic duration in their test sentence readings compared to SBE participants, while on the text setting rating tasks they showed a higher tolerance of stress mismatches.

General Discussion

The findings presented here provide evidence of the distinctive impact of the speech rhythm of an emerging variety of English—Multicultural London English as compared to standard Southern British English—on the melodic rhythms, text settings, and text setting intuitions of native speakers. The findings suggest that differences in the rhythmic characteristics of the two dialects are reflected in differences in the musical preferences of native speakers: MLE speech displays less variability in vocalic duration and syllabic prominence than SBE speech, and MLE speakers correspondingly produce melodic rhythms with less durational variability and are more tolerant of stress mismatches than SBE speakers. The findings support the suggestion (Dell & Halle, 2009) that the perceptual salience of stress distinctions in a language influences the degree to which native speakers tolerate stress mismatches in text settings (as already evidenced by studies of French and English speech and text settings; Lee & Todd, 2004; Temperley & Temperley, 2013), and provide further evidence that vocalic durational variability in speech is a factor, alongside other nonlinguistic factors such as stylistic preferences, influencing the durational variability of melodic rhythms produced by native speakers. They also provide evidence for the more general claim that tolerance for stress mismatches and preferences regarding melodic durational variability go together because of the link between vowel duration and prominence (Lee & Todd, 2004; Lee et al., 2014).

One question that arises from the study concerns the song corpora stress mismatch findings. As noted earlier, the rate of stress mismatching in the MLE songs is quite low when compared to the rates reported by Temperley and Temperley (2013) in their corpus of French songs. More particularly, it is also low when considered in light of the results from the experimental text setting rating tasks: most of the MLE participants had only a marginal preference overall for the stress-matched versions of the song fragments and a few actually had an overall preference for the stress-mismatched versions.

There are two explanations we can think of for the low rate of stress-mismatches found in the MLE songs. The first possibility is that the metrical analysis has resulted in an underestimate of the number of stress mismatches due to the general underlying assumption favoring stress-matched interpretations. The fact that a stress mismatch can be resolved by metrical reinterpretation does not by itself mean that the reinterpretation was intended by the songwriter or is inferred by the listener, and perhaps such reinterpretations may be less commonly intended/inferred by MLE than SBE songwriters/listeners precisely because they are more tolerant of stress mismatches. The second possibility is that the prevalence of stress mismatching in the MLE songs

is constrained by commercial factors. Given that stress mismatching—on the evidence presented here—seems to be disfavored by SBE speakers (and therefore probably also by speakers of other standard dialects of English with similar rhythmic characteristics), then the most commercially successful songs—with an audience comprising many non-MLE speakers—will tend to be those with fewer mismatches. These commercial constraints may also operate via the decisions of agents and managers before a song is released, through the choice of songs to promote as singles or include on albums, or at an even earlier stage in the songwriting process itself, through the influence of co-writers.

A further question concerns the performance of the SBE participants in the text setting rating tasks. Although they had a higher preference overall for the stress-matched fragments compared to their MLE counterparts, their results were more mixed than might be expected, with a few of them—as already noted—showing no evidence of any preference for the stress-matched fragments or indeed showing the reverse preference. It is possible that these different preferences reflect different listening habits, and hence are the result of stylistic rather than linguistic judgements, even if the stylistic character (mainstream pop/rock) of the experimental song fragments seems to us to make such judgements unlikely. However, we have no data from participants bearing on the question, and there is no evidence bearing on the more general question of which styles/genres might display high rates of stress mismatching (though genres originating in the Caribbean, such as reggae, would be plausible candidates given the rhythmic characteristics of Caribbean dialects such as Jamaican English; Wells, 1982). Alternatively, it could be that the results reflect changes currently underway in SBE itself: perhaps under the influence of a growing number of varieties of English like MLE that have emerged from the contact between ethnic and second-language varieties of the language (Nelson & Kang, 2015), SBE is

changing rhythmically, with a consequent reduction in the perceptual salience of rhythmic stress. However, in the absence of relevant research findings, the supposition for now remains speculative (for further discussion, see e.g., Crystal, 2003).

A final more general question concerns the extent to which the importance of lexical stress for spoken word recognition in a particular dialect/language affects native speakers' text setting preferences. There is evidence that native speakers of (probably standard American) English find stress-matched text settings more intelligible than stress-mismatched text settings (Gordon, Magne, & Large, 2011; Johnson, Huron, & Collister, 2014).[8] Is this true also for speakers of MLE and Spanish, and of other rhythmically similar dialects/languages with lexical stress (e.g., Italian), and if so to what extent does it reduce their tolerance of stress mismatching compared to speakers of languages such as French?

In conclusion, our study brings together findings from a diversity of sources (corpus analysis of popchart songs, text setting rating experiment, and analysis of participants' speech rhythm) to show the differential impact of MLE and SBE speech rhythm on the melodic rhythmic preferences and text setting preferences of native speakers. More generally, it provides further evidence of the influence of linguistic prosody on the rhythms and melodies of music.

Author Note

Correspondence concerning this article should be addressed to Daniel Müllensiefen, Department of Psychology, Goldsmiths, University of London, New Cross Road, London SE14 6NW. E-mail: d.mullensiefen @gold.ac.uk

References

Arvaniti, A. (2012). The usefulness of metrics in the quantification of rhythm. *Journal of Phonetics, 40*, 351-373.

Berg, T. (1999). Stress variation in British and American English. *World Englishes*, *18*(2), 123-143.

Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009 from http//: www.praat.org/

Carnegie Mellon University (2008). *The Carnegie Mellon University pronouncing dictionary* [v.07a]. Available at http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Cheshire, J., Kerswill, P., Fox, S., & Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of Multicultural London English. *Journal of Sociolinguistics, 15*, 151-196.

[8] No details are given of the linguistic background of the participants in either study other than that they were native speakers of English.

CRUTTENDEN, A. (2014). *Gimson's pronunciation of English* (8th ed.). New York: Routledge.

CRYSTAL, D. (2003). *English as a global language* (2nd ed.). Cambridge: Cambridge University Press.

CUTLER, A. (2005). Lexical stress. In D. B. Pisoni & E. Ramirez (Eds.), *The handbook of speech perception* (pp. 264-289). Oxford, UK: Blackwell.

DANIELE, J. R., & PATEL, A. D. (2004). The interplay of linguistic and historical influences on musical rhythm in different cultures. In S. Lipscomb, R. Ashley, R. O. Gjerdingen, & P. Webster (Eds.), *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 759-762). Adelaide, Australia: Causal Productions.

DANIELE, J. R., & PATEL, A. D. (2013). An empirical study of historical patterns in musical rhythm: Analysis of German & Italian classical music using the nPVI equation. *Music Perception, 31*, 10-18.

DAVIES, M. (2008). *The corpus of contemporary American English (COCA): 520 million words, 1990-present.* Available at http://corpus.byu.edu/coca/

DELL, F., & HALLE, J. (2009).Comparing musical text setting in French and in English songs. In J. L. Aroui & A. Arleo (Eds.), *Towards a typology of poetic forms* (pp. 63-78). Amsterdam: Elsevier.

GORDON, R. L., MAGNE, C. L., & LARGE, E. W. (2011). EEG correlates of song prosody: A new look at the relationship between linguistic and musical rhythm. *Frontiers in Psychology: Auditory Cognitive Neuroscience, 2*, 1-13.

GRABE, E., & LOW, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory Phonology 7* (pp. 515-546). Berlin, Germany: Mouton de Gruyter.

HANNON, E. (2009). Perceiving speech rhythm in music: Listeners classify instrumental songs according to language of origin. *Cognition, 111*, 404-410.

HANSEN, N. C., SADAKATA, M., & PEARCE, M. (2016). Non-linear changes in the rhythm of European art music: Quantitative support for historical musicology. *Music Perception, 33*, 414-431.

HAYES, B. (2009). Text setting as constraint conflict. In J. L. Aroui & A. Arleo (Eds.), *Towards a typology of poetic forms* (pp. 43-62). Amsterdam: Elsevier.

HAYES, B., & KAUN, A. (1996). The role of phonological phrasing in sung and chanted verse. *The Linguistic Review, 13,* 243-303.

HURON, D., & OLLEN, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception, 21*, 267-271.

JOHNSON, R., HURON, D., & COLLISTER, L. (2014). Music and lyrics interactions and their influence on recognition of sung words: An investigation of word frequency, rhyme, metric stress, vocal timbre, melisma, and repetition priming. *Empirical Musicology Review, 9*, 2-20.

KAISER, H. F., & RICE, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement, 34,* 111-117.

KOPF, D. (2015). *How many people take credit for writing a hit song?* Available at http://priceonomics.com/how-many-people-take-credit-for-writing-a-hit-song/

LEE C. S., KITAMURA C., BURNHAM D., & TODD, N. P. M. (2014). On the rhythm of infant- versus adult-directed speech in Australian English. *Journal of the Acoustical Society of America, 136*, 357-365.

LEE, C. S., & TODD, N. P. M. (2004). Towards an auditory account of speech rhythm: Application of a model of the auditory 'primal sketch' to two multi-language corpora. *Cognition, 93*, 225-254.

LIBERMAN, M. (2007). *Rock syncopations: Stress shifts or poly-rhythms?* Available at http://itre.cis.upenn.edu/~myl/languagelog/archives/005154.html.

LIBERMAN, M. (2009). *Bembé, Attis, Orpheus.* Available at http://languagelog.ldc.upenn.edu/nll/?p=1420.

LITTLEFIELD, H. (2006). *Syntax and acquisition in the prepositional domain: Evidence from English for fine-grained syntactic categories* (Unpublished doctoral dissertation). Boston University.

LONGUET-HIGGINS, H. C., & LEE, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1,* 424-441.

McGOWAN, R. W., & LEVITT, A. G. (2011). A comparison of rhythm in English dialects and music. *Music Perception, 28,* 307-314.

NELSON, C. L., & KANG, S.-Y. (2015). Pronunciation and world Englishes. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 320-330). Chichester, West Sussex, UK: Wiley Blackwell.

PATEL, A. D., & DANIELE, J. R. (2003a). An empirical comparison of rhythm in language and music. *Cognition, 87*, B35-B45.

PATEL, A. D., & DANIELE, J. R. (2003b). Stress-timed vs. syllable-timed music? A comment on Huron and Ollen (2003). *Music Perception, 21*, 273-276.

PATEL A. D., IVERSEN, J. R., & ROSENBERG, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *Journal of the Acoustical Society of America, 119*, 3034-3047.

PRZEDLACKA, J., & ASHBY, M. (2009). Acoustic correlates of glottal articulations in Southern British English. In W. Lee & E. Zee (Eds.), *Proceedings of the 17th International Conference of Phonetic Sciences* (pp. 1642-1645). Hong Kong: ICPHS.

RAMUS, F. (2002). Acoustic correlates of linguistic rhythm: Perspectives. In B. Bel & I. Marlien (Eds.), *Proceedings Speech Prosody* (pp. 115-120). Aix-en-Provence: Laboratoire Parole et Langage.

RAMUS, F., NESPOR, M., & MEHLER, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*, 265-292.

RODRIGUEZ-VAZQUEZ, R. (2010). Text setting constraints: A comparative perspective. *Australian Journal of Linguistics, 30*, 19-34.

SADAKATA, M. (2006). *Ritme and Rizumu: Studies in music cognition* (Unpublished doctoral dissertation). Radboud University Nijmegen, The Netherlands.

SADAKATA, M., DESAIN, P., HONING, H., PATEL, A. D., & IVERSEN, J. R. (2004). A cross-cultural study of the rhythm in English and Japanese popular music. In *Proceedings of the International Symposium on Musical Acoustics* (pp. 41-44), Nara, Japan: ISMA.

SHIH, S. S. (2014). *Towards optimal rhythm* (Unpublished doctoral dissertation). Stanford University.

TEMPERLEY, D. (1999). Syncopation in rock: A perceptual perspective. *Popular Music, 18*, 19-40.

TEMPERLEY, D. (2001). *The cognition of basic musical structures*. Cambridge, MA: MIT Press.

TEMPERLEY, N., & TEMPERLEY, D. (2011). Music-language correlations and the "Scotch snap." *Music Perception, 29,* 51-63.

TEMPERLEY, N., & TEMPERLEY, D. (2013). Stress-meter alignment in French vocal music. *Journal of the Acoustical Society of America, 134*, 520-527.

TORGERSEN, E., & SZAKAY, A. (2012). An investigation of rhythm in London English. *Lingua, 122*, 822-840.

VANHANDEL, L. (2005). *Setting a menu to music: Prosody and melody in 19th-century art song* (Unpublished doctoral dissertation). Stanford University.

VANHANDEL, L., & SONG, T. (2010). The role of meter in compositional style in 19th century French and German art song. *Journal of New Music Research, 39*, 1-11.

WEIR, M. (2005). *Jazz singer's handbook: The artistry and mastery of singing jazz*. Los Angeles, CA : Alfred Music Publishing.

WELLS, J. C. (1982). *Accents of English*. Cambridge, UK: Cambridge University Press.

WELLS, J. C. (2008). *Longman pronunciation dictionary* (3rd ed.). Harlow, UK: Pearson Education Limited.

WIGET, L., WHITE, L., SCHUPPLER, B., GRENON, I., RAUCH, O., &, MATTYS, S. (2010). How stable are acoustic metrics of contrastive speech rhythm? *Journal of the Acoustical Society of America, 127*, 1559-1569.

Appendix A: Accent Identification

While SBE is not difficult to identify, due its familiarity as a standard accent of English (see e.g., Cruttenden, 2014), the same is not true of MLE. However, MLE displays a number of characteristic segmental phonetic features (distinguishing it from Cockney, which it closely resembles), in particular: 1) the monophthongization and fronting of the FACE and PRICE diphthongs; 2) the monophthongization of the MOUTH and GOAT diphthongs; 3) the backing of /k/ to [q] before non-high vowels (see Cheshire et al., 2011). We relied on these features to help select the MLE artists for the song corpora analysis, and targeted them in the test sentences for the experiment, as follows (syllables targeting key features in bold):

1. She had a big **house** in the centre of **town**.
2. If you **go** back **now**, you'll have plenty of **time**.
3. They **cost** a lot of money, so **no**-one **buys** them.
4. A **cou**ple of **days la**ter, his **face** was **fine**.

Appendix B: Metrical Analysis

The goal of the metrical analysis was to derive the most plausible underlying metrical representation of the melodic lines, in terms of which text settings would likely be judged by listeners as stress-matched or stress-mismatched. In our initial analysis, we allowed only for metrical displacements: the "syncopation shifts" described by Temperley (1999, 2001), involving the displacement of one or more events from the beats they occur on in the underlying representation onto earlier beats in the surface representation.[9] The circumstances in which events are interpreted as metrically displaced are outlined in general terms by Temperley,

[9] There is evidence in our song corpora and elsewhere of displacements in the opposite direction: i.e., onto *later* beats in the surface representation. Such displacements create gaps in the melodic line, analogous to hesitation disfluencies in speech, and involve the movement of a phrase-medial/final event onto a later beat because of the occurrence of a rest/tied note (any subsequent events in the phrase may also be displaced onto later beats as a result). The phenomenon has not been described in the literature to our knowledge, but both types of metrical displacement may be species of more general phrasing phenomena, widely attested in many popular vocal styles (see e.g., Weir, 2005) in which the singer delays words/phrases ("back phrasing") or brings them forward ("forward phrasing") to achieve a conversational effect (frequently, however, leading to what amounts to a recomposition of the original melody rather than simple metrical displacement). The phrase "the storm that's raging on" in Ed Sheeran's "Lego House" provides an example: a 16th note rest after "the" results in a 16th note delay to "storm" and all the following events except "on." However, we did not allow for such displacements in our analysis, as there were almost no examples among the events analyzed (the sole exception, "raging," from the phrase just cited was instead analyzed as a conventionally displaced first event followed by an undisplaced second event, while "the storm" from the same phrase was not included in the analysis due to a discrepancy between the score and performance regarding the rhythm of the prior segment ending with "the").

but he does not specify an effective procedure for analysis. We did not attempt to develop such a procedure here; instead we devised the following semi-formalized heuristic, based on his account, and guided by the general assumption that listeners prefer interpretations that yield stress matches:

(1) A note-onset syncopated event (i.e., an event on a weak beat followed by a rest or tied note on a following strong beat) is interpreted as metrically displaced from the strong beat if (i) the displaced interpretation resolves a stress mismatch (or if the event is notated as held through the syncopation, and the displaced interpretation does not result in a stress mismatch); and (ii) the displacement is a single beat at the tactus level or lower.[10]

(2) A series of consecutive events is interpreted as metrically displaced if: (i) the displaced interpretation resolves a stress mismatch; (ii) the displacement is a single beat at the tactus level or lower; and (iii) the interpretation does not entail an overlap in the underlying representation with any subsequent events.[11]

Examples of melodic lines containing events analyzed as metrically displaced in order to resolve stress mismatches (or to reflect the notation) are shown in Figure A, along with their reinterpretations.

While the heuristic yielded musically plausible analyses in most cases, there were a number of unsatisfactory analyses. In four cases, due to considerations of musical parallelism and harmony, we judged there were clear stress mismatches, even though the heuristic yielded stress-matched interpretations. We therefore reanalyzed them as follows (see Figure B): 1) note-onset syncopated events were analyzed as displaced, despite resultant stress mismatches, in McFly's "Five colours in her hair" ("a" in "I threw a house party") and Ms Dynamite's "Dy-na-mi-tee" ("-few" in "breaking my curfew") because of musical parallelism (in "Five colours in her hair," "-bo-" in the preceding parallel phrase "everybody wants to know her name" is clearly displaced, while in "Dy-na-mi-tee," there are several parallel phrases where the corresponding syllable is also clearly displaced, such as "night" in the immediately

preceding phrase "hangin' out all night"); 2) in "The A team," "to-" in "we'll fade out tonight" was also analyzed as displaced, despite a resultant stress mismatch, because the note it occurs on (E) is more consonant with the chord change on the syncopation (Dmaj9) than the previous chord (Bmin11); 3) in "Anything could happen," "into" in "we fell into the river" was not analyzed as displaced, despite yielding a stress mismatch, because parallel phrases ("don't know" in "so you don't know the secret" and "our names" in "to see our names were written") could not also be plausibly analyzed as displaced.

Several other cases also yielded unsatisfactory analyses, but for a different reason: they seem to involve polyrhythmic reinterpretations (Liberman, 2007, 2009) rather than metrical displacement. In all four cases (see Figure C, upper panel), stress mismatches seem to be resolved not through metrically displaced interpretations but through local metrical reinterpretations of the phrases containing the mismatches, triggered by repeating rhythmic patterns (bracketed) reinforced by word repetition ("holla," sequence A; "money," sequence B) or assonance ("-round a-," "-bout a," "thou-," sequence C; "sofa," "dose of," sequence D). We therefore modified the analytical procedure to allow for such reinterpretations, as follows:

(3) An event is interpreted as part of a polyrhythmic figure if: (i) the local metrical reinterpretation resolves a stress mismatch yielded by the dominant metre without resulting in further stress mismatches; and (ii) the figure is part of a repeating pattern that implies the local reinterpretation.

The metrical reinterpretations of the four passages according to the revised procedure are shown in the lower panel of Figure C (we have assumed that the bar-level metrical units are unchanged, though the assumption may well be incorrect).

The final analysis yielded a total of 17 unresolved stress mismatches (in addition to the 4 cases described above). In the case of Labrinth's "Let the sun shine" (see Figure D), the stress mismatch ("because") is unresolvable because 1) the stress-mismatched events are not separated by a single unit at any metrical level, so there is no possible metrically displaced interpretation; and 2) there is no possible polyrhythmic reinterpretation that would resolve the mismatch.

The other cases are similar, except that what rules out a metrically displaced interpretation in each case is the fact that the stress-mismatched events cannot be moved without also moving subsequent events, and moving the subsequent events would create further stress mismatches. Figure E shows all 16 cases; the later

---

[10] Temperley (1999, 2001) allows for note-onset syncopated events to be multiply displaced (i.e., syncopated at more than one metrical level) such that the overall displacement is not a single metrical unit at any level. However, there were no examples of such displacements in our corpus.

[11] This condition is a restatement of Temperley's (2001) "deep representation ordering rule."

(a)

I'm gon na pick up the? pie?_____ ces

(b)

My three words have two mean_____ ings

(c)

I knew the pan ic was o? ver?

(d)

Bo? ttles? pop? ing? off, be fore you know it

(a)

I'm gon na pick up the pie ces

(b)

My three words have two mean ings

(c)

I knew the pan ic was o ver

(d)

Bot tles pop ping off, be fore you know it

APPENDIX B FIGURE A. Melodic lines containing events analyzed as metrically displaced in order to resolve stress mismatches, or to reflect the notation (original lines, upper panel; metrical reinterpretations, lower panel; stress-mismatched syllables in the accompanying lyrics followed by question marks): (a) Ed Sheeran, "Lego house," note-onset syncopated "pie-" ("pieces") analyzed as displaced, thereby resolving a stress mismatch; (b) Ed Sheeran, "Lego house," note-onset syncopated "mean-" ("meanings"), analyzed as displaced, as it is notated as held through the syncopation; (c) Ellie Goulding, "Anything can happen," "over" analyzed as displaced, thereby resolving a stress mismatch; (d) Labrinth, "Earthquake," "bottles popping off" analyzed as displaced, thereby resolving two stress mismatches.

events that would constitute stress mismatches under a metrically displaced reinterpretation resolving the earlier stress mismatches are underlined.

Table A summarizes the results of the analyses for each song (the results of the nPVI analyses are also included).

APPENDIX B FIGURE B. Melodic lines containing events interpreted as stress-mismatched, despite the general assumption favoring stress-matched interpretations (original lines, upper panel; metrical reinterpretations, lower panel; stress-mismatched syllables in the accompanying lyrics followed by question marks): (a) McFly, "Five colours," "a" and "house" analyzed as displaced, despite resulting stress mismatch; (b) Ms Dynamite, "Dy-na-mi-tee," "-few" analyzed as displaced, in the same way as preceding "night," despite resulting stress mismatch; (c) Ed Sheeran, "The A team," "to-" analyzed as displaced, despite resulting stress mismatch; (d) Ellie Goulding, "Anything can happen," "river" analyzed as displaced, resolving a stress mismatch, but "into" analyzed as metrically undisplaced, leaving the stress mismatch unresolved.

APPENDIX B FIGURE C. Melodic lines containing events analyzable as part of a repeating polyrhythmic figure (upper panel; figures and repeats bracketed), and their reinterpretations (lower panel), with accompanying lyric (stress-mismatched syllables followed by question marks): (a) Jessie J, "Do it like a dude"; (b) Jessie J, "Pricetag"; (c) and (d) Ed Sheeran, "You need me, I don't need you."

APPENDIX B FIGURE D. Labrinth, "Let the sun shine," melodic line containing unresolvable stress mismatch (stress-mismatched syllables followed by question marks).

Appendix C: Analysis of Participants' Speech Rhythm

Participants' readings of the selected sentence ("they cost a lot of money, so no-one buys them") were low-pass filtered at 8 kHZ and rescaled to a standard mean RMS value (in the case of 6 MLE readings, noise reduction was first applied to remove background noise). They were then segmented into vocalic ($n = 310$) and consonantal ($n = 278$) intervals (uninterrupted stretches of uniquely vocalic or consonantal material), ignoring the connective "so," using *Praat* (Boersma & Weenink, 2009), and following the method in Ramus, Nespor, and Mehler (1999). From this analysis, the standard rhythm metrics, %V, ΔV, ΔC, VarcoV, VarcoC, nPVI-V, and rPVI-C were derived (Grabe & Low, 2002; Ramus, 2002; Wiget et al., 2010), along with the syllable rate (syllables per second). The variability in the loudness of the syllabic nuclei was also calculated, using the Praat loudness model, along with the variability in their mean RMS intensity, following the method in Lee et al. (2014): the mean RMS intensities, I (in dB), and maximum loudness levels, L_{max} (in phons) of all the segmentable syllabic nuclei were first determined (all the vocalic intervals, plus 7 consonantal syllabic nuclei across 7 of the SBE readings), and then measures of global and serial variability were derived (the standard deviations of the mean intensities and maximum loudness levels across an utterance, ΔI and ΔL_{max}, and the raw PVI equivalents, rPVI-I and rPVI-L_{max}).

The segmentation of the readings presented a number of issues. First, as it was impossible to determine the exact start or finish of voicing in many of the readings, the sentence-initial and sentence-final consonants were ignored in the analyses. Second, the boundary between the two syllables in the word "no-one" was impossible to determine in most readings, so the two vocalic nuclei and intervening glide were analyzed in all cases as a single vocalic interval. Third, most of the MLE speakers and a number of the SBE speakers realized the final /t/ of "lot" as a glottal articulation falling short of complete closure, a type of realization that is common in a number of varieties of English (Przedlacka & Ashby, 2011) and whose

segmentation is problematic due to the absence of waveform/spectrogram cues. Given that the main acoustic correlate of such articulations is a drop in harmonicity (the harmonics to noise ratio, HNR), we used the method for HNR estimation in *Praat* (using the standard settings) to determine segment boundaries, as follows: we first located the harmonicity minimum, and then took the maxima on either side as marking the segment boundaries.

A more general issue concerns the small size of our dataset (a single sentence) as a basis for dialect rhythm comparison, especially in the light of evidence that there is large within-language/dialect rhythmic variability across sentences/utterances (due to syllabic complexity differences), elicitation methods, and speakers (Arvaniti, 2012; Wiget et al., 2010). However, given that the two dialects compared here have the same syllable structure, syllabic complexity is not a factor and hence does not need to be controlled for by the use of a wide range of materials, while the same elicitation method (reading sentences) was used for both groups, and the number of speakers ($n = 32$) was quite large compared to previous studies.

There was a final issue concerning two potential confounding factors that could have called into question the validity of the results:

1. Although all the MLE speakers and most of the SBE speakers produced readings with identical patterns of consonantal and vocalic intervals, some of the SBE speakers produced slightly variant readings: eight of them produced no vowel corresponding to the indefinite article "a" (1 misreading "a lot" as "lots," and the other seven syllabifying the following consonant /l/), and one of the eight also produced an elided realisation of "lot of" as effectively monosyllabic with a single vocalic interval. These variant readings would have resulted in different scores on all the measures (particularly the standard rhythm metrics) compared to a canonical reading, and if the scores on the three critical measures that yielded significant intergroup differences, nPVI, ΔL_{max}, and rPVI-L_{max}, in particular were higher, the effect would be to inflate the differences between the SBE and MLE groups and

APPENDIX B FIGURE E. Melodic lines containing unresolvable stress mismatches (stress-mismatched syllables followed by question marks), as a metrically displaced reinterpretation would yield further stress mismatches (events underlined): (a) Dappy, "No regrets"; (b) Ellie Goulding, "Anything can happen"; (c) Jessie J, "Do it like a dude"; (d) Jessie J, "Domino"; (e) Katie B, "Crying for no reason"; (f) Katy B, "Katy on a mission"; (g) Labrinth, "Earthquake"; (h) Ms Dynamite, "Put him out"; (i) Plan B, "Stay too long"; (j) Will Young, "Your game."

APPENDIX B TABLE A. *Results of the Metrical and nPVI Analyses of the Melodic Lines of all 32 Songs in the MLE and SBE Song Corpora*

Artist (group)	Song	Events/ pairs	MD	Poly-rhythm	Stress matches (tactus)	Stress mismatches (tactus)	Text (stress mismatches in bold)	nPVI
Dappy (MLE)	No regrets	80/44	14	0	42 (33)	2 (1)	"a *reflection*"	32.6
Sophie Ellis-Bextor (SBE)	Move this mountain	34/17	21	0	17 (11)	0		34.4
	Murder on the dancefloor	16/8	4	0	8 (7)	0		49.5
Tom Fletcher (SBE)	Five colours	72/39	34	0	38 (6)	1 (0)	"*a house* party"	47.0
Ellie Goulding (SBE)	Anything could happen	62/34	16	0	31 (23)	3 (1)	"*into* the river"; "*cover* your eyes"; "*after* the war"	28.3
	Starry-eyed	25/16	6	0	16 (14)	0		26.0
	The writer	68/38	9	0	38 (28)	0		49.3
Jessie J (MLE)	Do it like a dude	69/37	5	3	36 (31)	1 (1)	"*under* the collar"	15.3
	Domino	79/41	6	0	40 (36)	1 (1)	"*into* the sunlight"	27.0
	Laserlight	61/33	20	0	33 (19)	0		26.9
	Price tag	73/42	2	6	42 (28)	0		32.7
Katy B (MLE)	Crying for no reason	74/38	14	0	37 (32)	1 (1)	"how *many* more days"	29.7
	Katy on a mission	40/21	0	0	19 (18)	2 (2)	"*into* the room"; "*into* the tune"	17.9
Labrinth (MLE)	Beneath your beautiful	52/26	22	0	26 (23)	0		44.3
	Earthquake	37/19	4	0	16 (15)	3 (3)	"*ladies* and gentlemen"; "this is *something*"; "the *shirts* and the ties"	18.3
	Let the sun shine	52/28	7	0	27 (20)	1 (1)	"it's all *because* the sun"	39.2
Ms Dynamite (MLE)	Dy-na-mi-tee	77/42	7	0	41 (35)	1 (0)	"breakin' my *curfew*"	67.0
	Put him out	71/37	7	0	35 (25)	2 (2)	"*even* taking care"; "*taking* him back"	25.1
Plan B (MLE)	Prayin'	60/31	10	0	31 (22)	0		37.3
	She said	20/10	2	0	10 (9)	0		50.1
	Stay too long	20/11	5	0	10 (8)	1 (0)	"something *to go* wrong"	23.5
	The recluse	23/12	2	0	12 (10)	0		12.9
Ed Sheeran (SBE)	Drunk	64/33	13	0	33 (30)	0		50.6
	Lego house	40/22	8	0	22 (16)	0		37.0
	The A team	59/33	21	0	32 (25)	1 (0)	"fade out *tonight*"	51.1
	You need me	120/67	7	7	67 (49)	0		27.1
Florence Welch (SBE)	Rabbit heart	48/27	19	0	27 (15)	0		31.7
	Shake it out	94/50	25	0	50 (35)	0		37.5
	Spectrum	45/23	10	0	23 (14)	0		43.9
Will Young (SBE)	Don't let me down	28/14	14	0	14 (13)	0		51.6
	Jealousy	53/28	16	0	28 (15)	0		40.3
	Your game	30/15	7	0	14 (11)	1 (1)	"*baby* stop treading"	42.0

Note: The table includes for each song: the total number of events/syllable pairs included in the analysis, the number of events metrically displaced (MD) or reanalyzed as part of a polyrhythmic figure, together with the number of stress matches and mismatches (tactus-level (mis)matches in parentheses), and the text of any mismatched portion of lyric (mismatched syllables highlighted in bold), and the nPVI.

hence undermine the validity of the results. However, the eight variant readings had lower scores on the three measures compared to the other SBE readings; hence, they did not inflate the differences between the two groups.

2. As already noted, noise reduction was applied to six of the MLE readings to reduce background noise, and this could have affected the loudness measurements, resulting in different scores on the loudness measures: if as a result the scores were lower, the effect would be to inflate the differences between the MLE and SBE groups and hence undermine the validity of the results. However, the six readings with noise reduction had higher scores on the two loudness measures; hence, they did not inflate the differences between the two groups.

Appendix D: Song Fragment Scores

Below are the Sibelius scores of the stress-matched versions of the 10 experimental song fragments; the stress-mismatched versions were created by shifting the first vocal phrase (preceding the eighth-note rest) an eighth-note later (see Figure 1 for an example).

Appendix E: Pre-processing of the Preference Ratings Data

The 10 song fragments used in the listening experiment were designed as naturalistic music stimuli with the aim of balancing control over the parameter of interest—the (mis)match of melodic stress pattern with metrical structure—with many key characteristics of real music examples (e.g., original compositional and lyrical ideas, stylistically appropriate harmonic accompaniment, naturalistic performance) to achieve a high degree of ecological validity. However, when using a set of newly created and highly naturalistic stimuli to construct a novel test, it is important to confirm that the set of stimuli form a homogeneous test set, i.e., that they possess internal validity and they are all measuring the same attribute—in this case, the degree of preference for (mis)matched text settings. This is of particular importance if participants' ratings across different items are to be averaged to form a single score. The homogeneity of the test set can be formally assessed by subjecting the data to a factor analysis or principal components analysis (PCA) and testing whether all stimulus items load highly on a single factor or component, i.e., whether the test is unidimensional and ratings across different items can be averaged. If items load on more

than one component then this can suggest that ratings are also systematically influenced by other musical or non-musical factors in addition to the target attribute of interest. In this case, it is common practice for the construction of tests or questionnaires to eliminate items that do not load highly on the first component.

We ran a principal component analysis (PCA) on the ratings data from the first task to establish the unidimensionality and the homogeneity of the stimulus test set. The input data to the PCA were the differences in rating for the stress-matched and stress-mismatched version of each song fragment. These differences were computed for each participant and each song fragment individually. An initial PCA using all 10 items indicated that the test set as a whole only possessed a measure of sampling adequacy (MSA) of 0.68, which according to Kaiser and Rice (1974) is mediocre. The MSA values for individual variables showed that four stimulus items (song fragments 3, 4, 9, and 10; see Appendix D) had values of < 0.60 ("miserable") whereas the other six had values between 0.67 and 0.81 and thus were suitable for factor analysis. For the full item set the Bartlett test of sphericity reached significance, $\chi^2 = 64.93$, $df = 45$, $p = .03$, indicating that in principle the stimulus set as a whole had a correlational structure. However, the first component of the PCA solution only extracted 29.5% of the variance and the first three components had an eigenvalue of > 1 with a ratio between first and second component of 1.8. In addition, there was no clear "elbow" visible in the scree plot of this PCA solution, all of which suggested that the set of stimuli could not be considered unidimensional. We found that the four items with an MSA value of < 0.60 had loadings of < 0.36 on the first component while all other items showed loadings between 0.52 and 0.82 on the first component. These four items were therefore excluded from the set of stimulus items and a second PCA was run. The new solution showed very clear improvements on all measures. The MSA for the set of six items was 0.77 and all individual MSA values had values > 0.69. The Bartlett test of sphericity was highly significant, $\chi^2 = 40.59$, $df = 15$, $p < .001$, and the first component explained 45.4% of the variance extracted. There was now a clear elbow visible in the scree plot: the first component had an eigenvalue of 2.7 while the second component had an eigenvalue of only 1.0. All six items had loadings of between 0.56 and 0.83 on the first component. Thus, all indicators suggested a clear unidimensionality of the reduced stimulus set; additionally, a Cronbach's alpha of 0.73 indicated an acceptable level of internal reliability. Therefore, we accepted these 6 items (song fragments 1, 2, 5, 6, 7, and 8; see Appendix D) as the final set of test stimuli.

If there's a reason

APPENDIX D FIGURE A.

Love is all around

Love is all a-round   and the   sum-mer's just be-gun

APPENDIX D FIGURE B.

Good morning darling

Lyrics: *Good morn - ing dar - ling just called to find out how you are*

APPENDIX D FIGURE C.

Night is falling

Lyrics (measures 6–7): *Night is fall-ing   and I'm so   far a - way from home*

APPENDIX D FIGURE D.

She's feeling sad

APPENDIX D FIGURE E.

Tell me if you're ready

APPENDIX D FIGURE F.

People telling me

APPENDIX D FIGURE G.

Waiting on the corner

APPENDIX D FIGURE H.

When there's nothing

APPENDIX D FIGURE I.

You're tired of life

The vocals enter at measure 6 with the lyrics "You're tired of life and you're all of se-ven-teen"