AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychomusicology: Music, Mind, and Brain

ISSN: 0275-3987

2021, Vol. 31, No. 3-4, 107–122
https://doi.org/10.1037/pmu0000278

The Effect of Timing on the Singer's Tone of Voice

Yke Paul Schotanus
Institute for Cultural Inquiry, Utrecht University

Several studies have shown that music can be used to express and induce specific emotions. Only a few, however, investigate interactive expressions such as dominance, submissiveness, and sincerity. In the current study, it is hypothesized that aligning phrase onsets with strong beats supports perceived sincerity. In 2 online listening experiments, 52 (M = 26.35; SD = 7.25) and 89 (M = 28.39; SD = 11.37) participants, respectively, listened to 27 sung sentences and rated 15 Likert scale items for each stimulus. In Experiment 1, the whole sentence was timed either early, on beat, or late, and in Experiment 2 only the timing of the last stressed syllable was varied. Both experiments show that on-beat phrases are perceived as relatively "right" (a combination of sincerity, naturalness, and convincingness, among other things). Experiment 2 also shows that early phrases support perceived urgency, whereas late phrases support perceived upsetness. These results suggest that a syncopated note can be related to a rest on a strong beat either following or preceding it. In addition, several aspects of melody turned out to affect these factors as well. The results can be related to various theories and indicate that perceived "authenticity" can be modified by using specific musical features.

Keywords: prosody, syncopation, dynamic attending, music and emotion, music and language

Music is hypothesized to affect the meaning of a song, for example, its emotional meaning. A wealth of studies have addressed the question as to whether pure emotions can intentionally be expressed in such a way that the listener perceives them (see Cespedes-Guevara & Eerola, 2018; Gabrielsson & Juslin, 1996; Juslin & Laukka, 2003; Koelsch, 2011; Scherer et al., 2017; for reviews). Some studies focus more on felt emotions instead of perceived ones (see Schubert, 2013; for a review). Yet, very few studies investigate musical features communicating emotional states that are not particularly expressed by the performer or felt by the listener, but imply an interaction between performer and listener or performer and message. These may involve Frijda's (1986; p. 25) "interactive expressions" and "mimics," or Huron's (2015) "ethological cues and signals," which mainly concentrate on antagonistic versus affinitive behavior. For example, as in speech (Ohala, 1984), high pitches have turned out to signal submissiveness (Huron et al., 2006) and sociability (Shanahan & Huron, 2014). One example of these "affects" is sincerity (which has been mentioned by Pattison [2015]). Other interactive emotional communication through music may involve a nagging or a calming quality, which is defined by the effect on the listener (Schotanus, 2020a, 2020b, Part 3), and social or psychological constructions (Warrenburg, 2020) such as the impression of authenticity, which is very important in the music industry (Auslander, 1999; Frith, 1981/2007). Authenticity and sincerity may not seem to be directly related to emotion, yet, if a singer is interpreted to be inauthentic or insincere, this will affect both the listener's interpretation of the emotional state of the singer and the listener's own emotional state, and consequently it will also affect the listener's ability to connect with the music. As a result, in popular music, perceived authenticity can be decisive for an artist's success (Brackett, 1995; Eckstein, 2010; Frith, 1987/2007).

According to several authors (Findeisen, 2017; Pattison, 2015), perceived authenticity can deliberately be manipulated using specific musical features. For example, Pattison (2015) stated that either on-beat or off-beat phrasing (i.e., starting the first stressed syllable of a linguistic phrase either on a strong or a weak beat) is a useful technique for manipulating affect. And indeed, both singers and composers frequently create off-beat phrase onsets (Temperley, 2001), which is often interpreted to change affect (see, e.g., Brackett, 1995, pp. 58–74; Burns, 2000). Pattison argued that off-beat phrases sound less stable, which makes the listener feel that there may be some subtext to the lyrics, for example, because the singer is upset or less sincere.

These presumptions are in line with several theories. First, the dynamic attending theory (DAT; Jones, 1976; Large & Jones, 1999) assumes that our attention oscillates in accordance with a given speech rhythm or musical rhythm. Consequently, in Western listeners, attention is optimal at strong beats, and in a stress-timed

This article was published Online First August 26, 2021.

Yke Paul Schotanus https://orcid.org/0000-0001-9992-4366

This research was supported by the Netherlands Organization for Scientific Research: 023.004.078.

Parts of this study have previously been published in my doctoral dissertation (Schotanus, 2020b, pp. 232–246) and were partly based on Schotanus (2018). Experiment II was not reported on before. I thank my supervisors Dr. E. G. J. Wennekes, Dr. F. Hakemulder, and Dr. R. H. Willems for their support and comments. I have no conflict of interest to disclose.

Correspondence concerning this article should be addressed to Yke Paul Schotanus, Institute for Cultural Inquiry, Utrecht University, Muntstraat 2a, 3512 EV Utrecht, the Netherlands. Email: schotschrift.teksten@planet.nl

language, such as Dutch, it is optimal at stressed syllables. In line with this, several studies have shown that stressed syllables presented off-beat were more difficult to process than stressed syllables presented on-beat (Gordon et al., 2011; Quené & Port, 2005). Furthermore, off-beat phrases can cause loud rests (London, 1993), which are known to cause substantial brain activity (Ladinig et al., 2009) and may distract from language processing. It is not sure, however, whether on-beat silences in the voice part are indeed perceived as loud rests if the accompaniment continues. Schotanus (2020b, Chapter 3.2) did not even find on-beat silences in a cappella songs to be perceived as loud rest. However, Temperley (2009) has found that in polyphonic music an onset in a metrically weak position is more likely to be succeeded by an onset in a metrically strong position in the same stream than in a different stream or by a rest, and Witek, Clarke, Kringelbach, & Vuust, et al. (2014) have shown that syncopations of one stream within a polyphonic stimulus destabilize a rhythmic pattern even if events in other sound streams occur on-beat. So, both dynamic attending and loud-rest-processing can affect the processing of words and create a feeling of disbalance in off-beat phrases. On the other hand, stressed syllables occurring off-beat may be also perceived as accentuated, as major musical events occurring off-beat are perceived as relatively loud (Bouwer & Honing, 2015).

Yet, there is more to it. The musical foregrounding hypothesis (Schotanus, 2015, 2020b) states that obstructions of language processing caused by musical events such as off-beat phrases are, if possible, interpreted as meaningful prosodic cues. As a result, the timing of the stimulus, or the listener's confusion about it, can be attributed to the intentions or the emotional state of the singer. Therefore, early onsets are hypothesized to be perceived as hasty or compelling, and late onsets as lingering, or uncertain, or something similar, whereas on-beat phrasings will be perceived as relatively natural, convincing, and at ease. On the other hand, however, as argued before, off-beat phrases can also increase the salience of either words or melody (Huron, 2016, p. 98). On top of that they can be hypothesized to induce more positive emotions, as rhythmical irregularities within a regular pattern can heighten the subcortical reward for correctly predicting the beat (Menon & Levitin, 2005), and medium degrees of syncopation support pleasure and wanting to move (Witek, Clarke, Wallentin, et al., 2014).

Building on an earlier pilot experiment (Schotanus, 2018), these assumptions were tested in two online listening experiments, with on-beat phrases, early phrases (one quarter or one eighth before the down beat), and late phrases (one quarter or one eight after the beat). Although syncopations usually are considered to be musical events occurring on a weak beat followed by a rest on a strong beat (Condit-Schultz, 2020; Koops et al., 2015; Longuet-Higgins & Lee, 1984; Tan et al., 2019; Temperley, 2019), in the current study it is hypothesized that syncopations can also be musical events occurring on weak beats preceded by a rest on a strong beat. In these cases, the sense of syncopation will be caused by a combination of phrase structure, musical structure, and DAT.

Pilot Study

In the pilot experiment (Schotanus, 2018), 30 participants were presented with a series of short sung sentences, accompanied on the piano, and after each sentence they were asked to answer seven questions (i.e., Likert-scale items) concerning the stability, emotionality, and aesthetic quality of the music, the lyrics or the fragment (see Table 2), and three questions concerning the singers sincerity, insecurity and compellingness. To mask the aim of the experiment, only these three items explicitly addressed the singer's tone of voice. All sentences consisted of either five or six syllables, three of which were stressed ones.

The results of this pilot experiment indeed indicated that ratings for several aspects of aesthetic and psychological stability (sincerity, naturalness, interesting melody, and good match between music and lyrics) were correlated and merged into one factor. This factor was called "Rightness" and not, for example, Positive Value, because "Rightness" has both aesthetic and moral implications. Furthermore, Aaftink (2014) also recognized something being "just right" as an important value in aesthetics, be it without the moral implication of sincerity. As expected, Rightness seemed to show a close to significant (p = .050) main effect of timing, but only if an insignificant interaction with meter was involved, indicating that trochaic sentences, all sung to melodies without a pick-up note, are rated less right in on-beat versions.

Apparently, the pick-up note in iambic sentences plays a substantial part in the effect of condition. One explanation may be that the pick-up note accentuates the beat in the melody (which would be in line with the hypotheses of Hansen et al. [2018]) and thus also accentuates alignments and misalignments with the accompaniment. Another explanation may be that in late versions the pick-up notes are dissonant with the accompaniment. Yet, these dissonant sounds do not occur in each sentence. Furthermore, early versions cause dissonance in both iambic and trochaic sentences, but there is no difference in Rightness ratings between early and late ones. On top of that, in popular music it is quite common that melodies are syncopated whereas the accompaniment is not (Burns, 2000; Temperley, 2001, p. 239–247), and that these syncopations are perceived as anticipatory syncopations (i.e., syncopations anticipating the strong beat; Tan et al., 2019). It is therefore more likely that the main effect of timing is due to dynamic attending and/or loud rests.

Unfortunately, in this pilot study the effect of Rightness was only close to significant, and the effects of the other factors, called "Compellingness" (mainly a combination of compellingness and emotional load) and "Upsetness" (mainly a combination of being insecure, nonenergetic, and emotional), were not significant at all. Yet, the poor results of the experiment do not allow for the conclusion that the effect of timing on perceived tone of voice is either absent or negligible. Several confounding factors may have weakened the results. First, the fact that only a few of the Likert-scale items addressed the singer's tone of voice seemed to have hampered the clarity of the factors emerging from a factor analysis of the data. Second, in the stimuli, the complete melodic line was shifted in relation to the accompaniment. Consequently, all three stressed syllables in a sentence occurred either on a strong or a weak beat. Gordon et al. (2011) found that beat tracking shifts from the rhythm of strong beats to the rhythm of stressed syllables when stressed syllables consistently occur on weak beats.

Other factors that may have weakened the results are the relatively small sample size (30 participants), and the relatively large but unbalanced variety in linguistic and musical properties which also could affect the perceived tone of voice. Several of these properties indeed turned out to do so. First, tempo (or note rate) seems to affect Compellingness and Upsetness in opposite directions, probably as an indicator of energy. Second, meter (iamb or trochee) seems to affect Compellingness and Rightness in opposite directions. Third, a three-four time decreases both Rightness and Upsetness. Fourth, nonclosure increases Compellingness but decreases Upsetness. And finally, melody seemed to affect the emotional meaning of a song and the interpretation of the tone of voice of the singer more strongly than sentence type. Only Compellingness is slightly enhanced by commands and requests, compared with questions, statements, and ellipses.

Current Study

In the current study, the main methodological flaws of the pilot study were addressed stepwise. First, in Experiment 1, the number of Likert-scale items was increased, and their nature was more clearly hypothesis related, that is, more clearly related to stability and convincingness, to hesitance and upsetness, or to a pressing quality and compellingness. In addition a few items related to positive feelings were added, in order to distinguish between rightness and positive feelings.

Second, in Experiment 2, the stimuli were also altered. This was done in such a way that the first and the second stressed syllable of a sentence were always on beat, and that only the third one was either early, on beat, or late. In addition, the number of stimulus versions was increased from three to five (very early, early, on beat, late, and very late) in the second experiment, and the number of participants in both.

Although the main focus of these experiments is on the effect of timing on tone of voice, irrespective of text and music, additional analyses will explore the effect of linguistic meter, tempo, measure, closure and sentence type, as such analyses have turned out to be more powerful than regressions with a random intercept for melody, or with melody as a fixed factor, in the pilot experiment. These analyses should be interpreted with caution, because neither the aspects of melody nor the categories of sentence type are sufficiently counterbalanced. However, they will be interesting in an exploratory way, as they give indications as to why certain melodies make a singer sound more sincere, more upset, or more compelling.

Experiment 1

The results of a pilot experiment only showed a marginal effect of timing (only on Rightness, $p = .05$, and only when interacting with meter). This raised the question whether the insignificance of the effect of condition was due to one of several confounding factors, one of which was a lack of clarity in the factors, caused by the nature of the Likert-scale items participants had to rate. A second listening experiment was therefore created with exactly the same stimuli but different Likert-scale items to represent the stimuli. The aim was to choose items that cluster more clearly in aesthetic and psychological stability (a bit like Rightness in the pilot study), aggressive instability (a bit like Compellingness), and weak instability (a bit like Upsetness). Whereas on-beat phrases were hypothesized to be associated with Rightness, early ones were expected to be associated with aggressive instability because they were assumed to have a pressing quality, and late ones with weak instability because they were assumed to be associated with hesitance. However, such instability does not have to be interpreted negatively. Therefore, off beat phrasings were also expected to be associated with positive feelings such as an impression of friendliness. Apart from timing, several other stimuli-properties where assumed to affect the ratings. An additional exploratory inspection of these effects was reported on briefly.

Method

Participants

A total of 52 participants aged between 18 and 48 ($M = 26.35$; $SD = 7.25$; 30.8% female) completed the survey and were accepted. Participants were recruited via the research platform Prolific Academic, with the only constraint that they had to speak Dutch (see Schotanus, 2020b, p. 381, for the invitation text and the survey introduction). Trials were rejected if the answer to the final control question was not correct, the participant had taken less than 15 min, or the ratings followed a suspicious pattern (e.g., if large numbers of subsequent items were rated 1). Accepted trials were remunerated.

Procedure

At a time and place chosen by themselves, participants listened to 27 sung sentences, preceded and supported by a piano accompaniment (total track duration about 12 s). After each sentence, they rated on a 7-point Likert scale whether they agreed with the following statements: the singer sounded sincere, convincing, hesitant, indifferent, restless, or greedy; the whole fragment sounded stable, cheerful, hasty or insecure; the voice sounded nagging or upset; the lyrics sounded friendly or compelling, and whether lyrics and music were a good match. The last item and the one on stability which are not necessarily related to an affect, were included in order to investigate whether musical stability is indeed associated with sincerity or convincingness. Several questions address the lyrics, the music, or the whole instead of the singer, to mask the aim of the experiment. A factor analysis was planned to unveil the connections between the ratings.

Halfway through the experiment, participants answered a few questions about their musical and literary training, using a Dutch translation of the Gold MSI Musical Training subscale (Bouwer et al., 2021; Müllensiefen et al., 2014), and a five-item questionnaire concerning literary training presented in Schotanus (2020b, Chapter 2.1). After a factor analysis on the latter, only one factor could be retained: Writing experience[1]. Musical training and Writing experience were used as covariates in the regressions.

Stimuli

A total of 27 sentences were sung to nine melodies. There were 13 uses of the imperative, five questions, six statements, and three elliptical sentences (e.g., an address, or: "Door red, shutters green"; see Appendix A for a complete overview, and see Schotanus, 2019, for all MP3 files). All sentences consisted of three metrical feet; 18 consisted of trochees, nine of iambs. Consequently, the former consisted of five syllables and the latter of six. The melodies, created by me, were meant to express feelings appropriate to at least one of the sentences sung to it. Furthermore, they were meant to vary in melodic contour, key, tempo, and harmony, as off-beat phrasing was hypothesized to have a general effect independently of text and music. The melodies also varied in measure: Six melodies were in three-four time, three in four-four time (for an impression of the distribution of these properties over the different melodies, see Table 1; for the

[1] Kaiser-Meyer-Olkin Statistic (KMO) = .724; determinant: .504. Factor loadings: I often write academic and journalistic texts: .65; I often write creative texts: .63; I like to play with words: .51; Hours spent writing: .65. One variable, "I do not mind the wording of what I say" did not meet the criteria for inclusion Measure of Sampling Adequacy (MSA $< .5$).

Table 1
Mean Nonclosure Rating, Meter, Measure, Tempo, and Sentence Type Per Melody

Melody	Nonclosure	Meter	Tempo	Measure	Sentence type[a]
100	0.13	Trochee	Low	4/4	i/s/s
130	0.22	Trochee	Moderate	3/4	i/i/i
150	0.48	Trochee	Low	3/4	i/s/q
170	0.10	Trochee	Moderate	4/4	i/i/q
190	−0.28	Trochee	Low	3/4	i/i/q
210	−0.69	Trochee	Low	3/4	e/e/s
220	0.44	Iamb	High	3/4	i/q/s
250	−0.28	Iamb	High	4/4	i/i/i
280	−0.13	Iamb	Moderate	3/4	q/s/e

Note. Nonclosure ratings were largely in line with musical theory. A fragment ending on the dominant (220) received high nonclosure ratings and a fragment in which both melody and accompaniment ended on a tonic received the lowest (210). However, surprisingly, fragment 190, with the chord progression Em-Em-Em(add2)-Em(add2)-Em-A-F#/A#, ending with an out-of-key chord, also received a very low nonclosure rating, even from musically trained listeners.
[a] Sentence type: i = sentence in imperative mood; s = statement; q = question, e = ellipsis.

scores of the stimuli, see Appendix B). The sung sentences were preceded and accompanied by piano music improvised by Christan Grotenbreg. He was asked to create different kinds of accompaniments, whether or not using the harmonies suggested by me, but always establishing a beat. As a result, all sentences would have a clear and similar rhythmic structure, aligned to a well-established beat, but would sound relatively interesting and ecologically valid where possible, given the atomic design of the study. All piano parts were recorded only once. The sentences were sung with the piano part on the headphone.

In most cases, neither the melody nor the accompaniment ended on a tonic (see Table 1). To measure harmonic closure for all the melodies, a separate experiment was run among 40 participants, recruited via Amazon Mechanical Turk, who did not speak Dutch and did not understand the content of the sentences. They were presented with an example of each melody and were asked to rate whether this musical fragment sounded as if it had finished, whether there was some remaining musical tension after the last note, and whether in their minds they heard some final notes they would expect to follow that last one. A principal components factor analysis resulted in one factor representing nonclosure (see Table 1). Several participants were Indian, but their ratings did not differ significantly from those of the American participants.

All sentences were sung once, but they were digitally edited in three different ways: early, on beat, and late (see Figure 1). In the on-beat version, the three stressed syllables in the sentence were aligned with the first beat, in the early versions the onsets of the stressed syllables were timed one-eighth note before the first beat, and in the late versions they were aligned with the second beat. In most but not all cases, the on-beat version was the original one. Of course it can be a confounding factor that most of the off-beat versions are digitally altered, whereas most of the on-beat versions are not, but singing off-beat deliberately is very hard to do in a neutral way. Of course, off-beat phrasing is quite usual in popular music, but it is very likely that, whether consciously or not, singers will use performance tricks in order to enhance the effect of timing, or that singing off-beat requires some physical or psychological effort which will inevitably affect the performance prosodically. Therefore, singing part of the stimuli off-beat would add other confounding factors.

No fillers were created, as it is impossible not to time either off beat or on beat. Afterward, the 81 musical fragments were distributed over three sets of stimuli, in such a way that each participant heard each sentence and each melody once in each version.

All sentences were sung by me and recorded by Christan Grotenbreg in his studio. The piano intros were improvised by Christan Grotenbreg on a keyboard connected to ProTools 10 (Desktop recording). The voice was recorded using a Neumann TLM 103 microphone, and an Avalon VT 737 SM amplifier. Digital conversions were conducted using Apogee Rosetta. To avoid confusion concerning purity and timing, voice-treatment software was used: Waves Tune, Renaissance Vox compression, and Oxford Eq.

Analysis

The ratings were analyzed using principal axis factoring with rotation (direct oblimin). Subsequently, crossed classified regression analyses were conducted on the factors using mixed models.

Results

Factor Analysis

Principal axis factoring yielded four factors with eigenvalues larger than 1, three of which indeed seemed to be clearer versions of the three factors in the pilot experiment (see Table 2). As predicted, one of these factors, showed a close relationship between musical stability, convincingness, and sincerity. As in the pilot experiment, this factor is called Rightness because of the implications of moral rightness and the reference to Aaftink (2014). Admittedly, clear indications of positive valence, such as cheerfulness and friendliness, contribute to this factor as well, but they do so more powerfully to a separate factor also associated with musical stability: Pleasantness. As intended, two other factors, Upsetness and (the absence of) Urgency, appear to be improved versions of the factors Upsetness and Compellingness in the pilot experiment. As both Upsetness and Urgency reflect negative qualities, it is clear that neither of them reflect negative valence as such, which is in line with the hypothesis that they would reflect affects such as weakness and aggressiveness.

An investigation of the means per factor per condition (see Figure 2) indicates that on-beat phrases support Rightness, and to a lesser extent Pleasantness, whereas off-beat phrases support Urgency and late phrases support Upsetness.

Figure 1
Stimulus Example

Note. One sentence: "Liefste, liefste, blijf" (Darlin', Darlin', stay.) in Three Versions: Early (A100); On Beat (A100a); and Late (A100b).

Regressions

To investigate whether the differences were significant, several series of regressions were run on all factors. First, crossed classified mixed model regressions with random intercepts for participant, sentence and melody, and condition, musical training, writing experience, and disinterest in wording as fixed effects. Second, crossed classified mixed model regressions with random intercepts for participant and sentence, and with timing, musical training, writing experience, sentence type, measure, tempo, meter, and nonclosure as fixed effects. All variables that did not show a significant effect were deleted, except condition. Finally, the models found in the second series were implemented in regressions with the intercepts of the first series, which in one case improved the model. Table 3 shows the results.

As hypothesized, on-beat phrasing now turned out to support Rightness ($p < .01$); however, the effect of timing on the other factors was not significant. This may be due to the fact that in early and late sentence versions the whole sentence was shifted in relation to

Table 2
Factor Analysis of Affect Ratings, Experiment 1

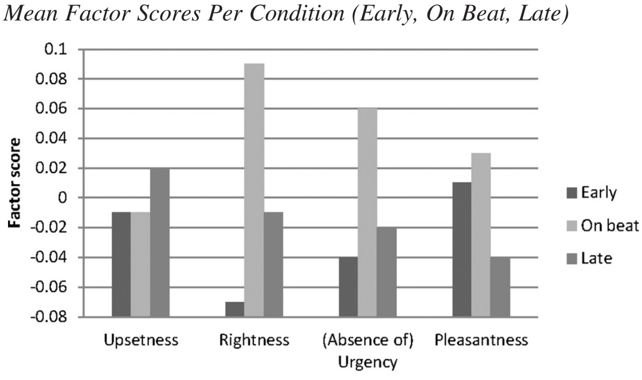
Items	Factors			
	Upset	Right	Pressing	Pleasant
All sounds insecure.	**.750**	−.238	−.225	
The voice sounds upset.	**.744**		−.296	
The singer sounds restless.	**.706**	−.213	**−.441**	−.186
The singer sounds hesitant.	**.630**	−.168	−.151	.233
The singer sounds sincere.	−.151	**.772**		.101
The singer sounds convincing.	−.290	**.751**	−.208	.133
Lyrics and music are a good match.		**.597**		.230
All sounds stable.	−.333	**.549**		**.401**
The voice sounds nagging.	.305		**−.805**	
The lyrics sound compelling.		.136	**−.690**	−.131
The singer sounds greedy.	.387		**−.644**	.241
All sounds hasty.	.290	−.101	**−.551**	.167
All sounds cheerful.	−.257	**.468**		**.695**
The lyrics sound friendly.	−.158	**.484**	.248	**.653**
The singer sounds indifferent.	.321	−.236	−.271	.329
Initial eigenvalue	3.96	2.80	1.67	1.32
Percentage of variance predicted	26.37	18.69	11.13	8.79
Rotated sum of squared loadings	2.76	2.50	2.38	1.46

Note. Factor loadings $<.1$ are deleted, factor loadings $>.4$ are in bold. Extraction method: Principal axis factoring. Rotation method: Oblimin with Kaiser Normalization. Kaiser-Meyer-Olkin Statistic (KMO) = .806; determinant = .004; Measures of Sampling Adequacy (MSA) for all variables $>.5$.

the accompaniment, or to one of the other causes suggested in the introduction. In the case of Pleasantness, the lack of significance may also be due to the small number of questionnaire items contributing to this factor. Additional items such as "The message sounds warm," could strengthen this factor and increase significance of an effect of timing. On the other hand, the effect of timing on Pleasantness may also be somewhat ambiguous. Despite the connection with musical stability, Pleasantness is supposed to be associated with off-beat phrases because of their emotional subtext and the moderate syncopations (Witek, Clarke, Wallentin, et al., 2014).

For most factors, the models without a random intercept for melody were more powerful than those with a random intercept for melody. Only the model for Upsetness was improved by adding the random intercept for melody. However, changing the models hardly affected the effect of timing.

Figure 2
Mean Factor Scores Per Condition (Early, On Beat, Late)



An exploratory investigation of these alternative models, showed that they were largely in line with those in the pilot experiment. Upsetness was related to a relatively slow tempo ($p < .001$). A three-four time measure and a trochee decreased Rightness ($p < .001$ and $p < .01$, respectively), and a three-four time measure, a trochee, and a high tempo increased the sense of Urgency ($p < .01$, $p < .01$, and $p < .001$, respectively). The new factor Pleasantness turned out to be supported by an iambic meter ($p < .001$).

A difference with the pilot experiment is that in Experiment 1 Rightness showed a significant effect of sentence type ($p < .05$), indicating that statements support Rightness, whereas the effect on Urgency is only close to significant ($p < .06$).

Remarkably, in this experiment, there were no significant effects of musical training, writing experience, or nonclosure. In the case of writing experience, this may be due to a ceiling effect. People who speak Dutch and are also active on Prolific Academic may have a relatively high educational level and have relatively extensive writing experience.

Conclusion

As expected, changing the set of affect ratings used in the pilot experiment resulted in clearer factors, and subsequently, clearer regression results. However, the results may still be attenuated by the fact that the timing differences complied to complete sentences, not parts of them.

Experiment 2

In Experiment 1, the results of the pilot experiment turned out to be improved by changing the set of Likert-scale ratings for each stimulus to get clearer factors emerging from the data. In the current experiment the design will be further improved by using

Table 3
Crossed Classified Mixed Model Regressions on Upsetness, Rightness, the Inverse of Urgency, and Pleasantness

	AIC interaction	AIC model	F_{timing}	F_{tempo}	Early	On beat	Late
With melody							
Upsetness	3,218.71	3,221.82	0.54		−.02 (.16)	−.01 (.16)	.03 (.16)
		3,217.40	0.54	9.41*tp			
Rightness	3,503.39	3,496.38	5.28**		−.07 (.12)	.09 (.12)	−.02 (.12)
Urgency	3,091.21	3,091.53	1.92		−.03 (.16)	.05 (.16)	−.02 (.16)
Pleasantness	3,187.67	3,187.10	2.45[a]		.04 (.17)	.02 (.17)	−.06 (.17)
Without melody							
	AIC interaction	AIC model	F_{timing}	F_{tempo}	F_{stype}	$F_{measure}$	F_{meter}
Upsetness	3,229.94	3,220.79	0.57	18.38***			
Rightness[b]	3,506.02	3,482.19	5.19**		3.66*	17.91***	13.65**
Urgency	3,095.47	3,080.71	1.89	16.94***	2.82^{+}	9.89**	10.41**
Pleasantness	3,195.68	3,178.06	2.48[a]				26.76***
Random intercepts for best performing models							
	Participant		Sentence		Melody		
	Est (*SD*)	*Z*	*Est* (*SD*)	*Z*	*Est* (*SD*)	*Z*	
Upsetness	0.31 (0.06)	4.80***	0.06 (0.02)	2.58*	0.06 (0.04)	1.5	
Rightness[b]	0.31 (0.06)	4.75***	0.02 (0.01)	2.25*	-		
Urgency	0.32 (0.07)	4.85***	0.08 (0.02)	3.29**	-		
Pleasantness	0.25 (0.05)	4.75***	0.13 (0.04)	3.42**	-		

[a] $p < .85$. [b] Akaike's Information Criterion (AIC) for a model including c*syllable count was 3,481.69, although c*syllable count was not significant; in this case pairwise comparisons for both early-on beat*** and on beat-late* were significant; in other regressions only the difference early-on beat** is significant.
* $p < .05$. ** $p < .01$. *** $p < .001$. $^{+}$ $p < .06$.

different stimuli. Instead of the whole sentence, only the last stressed syllable was phrased early, on beat, or late. This is hypothesized to enhance the effect of timing for several reasons. First, the irregular pattern will prevent the listeners from shifting their attention from the rhythm of the music to the rhythm of the language (Gordon et al., 2011). Second, the rhythm will be established more clearly. And third, whenever the language allows for such an interpretation, the deviant timing of the last word is hypothesized to be perceived as a prosodic accent. In most sentences, this is indeed the case as the last word is often very important for the meaning of the sentence. Most sentences are not even complete without the last word (see Appendix A). Hence, there is also no risk that the effect of timing will be diminished because people do not expect another word.

To improve the results further, the number of versions per sentence was increased. In the previous experiment, the off-beat timing was either one eighth early or a quarter late. In the current experiment, off-beat timings of one quarter early and one eighth late were added. Words that occurred a quarter early were supposed to occur too early to be perceived as sung to a syncopated note, as opposed to words that occurred an eighth early (Tan et al., 2019). On the other hand, words occurring a quarter early would occur relatively fast in relation to the preceding words. This can affect the tone of voice in another way. Among the late versions, the timing cannot affect the loudness of the loud rest, as this loudness is supposed to be dependent on the last note before the loud rest (Ladinig et al., 2009; Longuet-Higgins & Lee, 1984). Yet, if the absence of a musical event is perceived as a musical event (Ladinig et al., 2009), the duration of this event is likely to affect its saliency, just as it accentuates other musical events (Huron & Royal, 1996). The effect of a quarter delay is therefore supposed to be stronger than that of an eighth.

Method

Participants and Procedure

Eighty-nine participants completed at least the first part of the survey and were accepted. They were aged between 18 and 74 (M = 28.39; SD = 11.37; 31 female, 58 male; 78 native speakers of Dutch) and predominantly nonmusicians ($M_{\text{GoldMSI MT}}$ = 18.51; SD = 8.73). As was the case in Experiment 1, all participants were recruited via Prolific Academic, completed the survey online at their own time and place, were informed that by participating in the survey they gave me permission to use their answers for my research, and were remunerated. They were randomly assigned to one of five sets of stimuli and questions. Design and wording of the questionnaires were largely the same as in Experiment 1, except that two extra attention control questions were added, that the survey had to be split into two (because with five versions it became too large), and that the stimuli were slightly different. Again, a factor analysis of the five items concerning writing experience yielded just one factor with an eigenvalue > 1, Writing Experience[2].

Due to a survey error, only 34 of the participants (hereafter: the small selection [SS]) were able to complete a correct version of both parts of the survey and include their prolific ID in either part; these people were aged between 18 and 64 (M = 28.47; SD = 10.63; 13 female, 21 male; 31 native speakers of Dutch; $M_{\text{GoldMSI MT}}$ = 16.82; SD = 8.68).

Another 34 people were able to complete a correct version of both parts of the survey but did not mention their prolific ID in the first part, so the two parts could not be matched. The SS was

[2] KMO = .604; determinant: .657. Factor loadings: I often write academic and journalistic texts: .52; I often write creative texts: .48; Hours spent writing: .87. The other variables did not meet the criteria for inclusion (Measures of Sampling Adequacy (MSA) < .5).

therefore expanded with a combination of these data concerning the second part of the survey and the results of 34 participants concerning the first part of the survey, selected randomly out of the 55 remaining sets of answers (hereafter: the large selection [LS]). Hence, the group of participants who have reacted to the first 18 sung sentences is slightly different from the group of participants who have reacted to the last 11, but only the characteristics of the first group can be specified. These 68 participants were aged between 18 and 74 ($M = 29.23$; $SD = 12.40$; 25 female, 43 male, 61 native speakers of Dutch; $M_{\text{GoldMSI MT}} = 17.82$; $SD = 8.57$).

Stimuli

The same 27 sung sentences were used as in Experiment 1, in the same order and with the same accompaniment. However, the two off-beat versions were replaced by four newly created ones. In each case, the last metrical foot of the sentence was moved an eighth or a quarter forward or backward in relation to the accompaniment (see Figure 3). Again, these sentence versions were created by digitally editing the original recordings to avoid effects of singing or pronunciation. The sentence versions were distributed as balanced as possible over five survey versions (see Appendix A).

Analyses

As in Experiment 1, the ratings were analyzed using principal axis factoring with rotation (direct oblimin). Subsequently, mixed model regressions were conducted, with timing (i.e., sentence version) as a fixed factor, GoldMSI MT and writing experience as covariates (only in regressions within SS), and random intercepts for participant, sentence, and melody. Additional explorative regressions were conducted with timing, sentence type, linguistic meter, tempo, measure, and nonclosure as fixed factors, and random intercepts for participant and sentence.

Results and Discussion

After a principal axis factoring analysis of the stimuli ratings, four factors with eigenvalues larger than 1 were retained: (the inverse of) Pleasantness, Urgency, Upsetness, and (the inverse of) Rightness (or Stability; see Table 4). Regarding the factor loadings, these factors are by and large the same as those in Experiment 1. For rhetorical clearness, the factors inverse of Rightness and inverse of Pleasantness were multiplied by -1 to create the factors Rightness and Pleasantness. As expected, these factors are quite similar to those in Experiment 1, which suggests that they indeed represent realistic psychological phenomena.

Regrettably, a further analysis of the effect of timing on factor scores was complicated, due to unforeseen problems with the survey software. As a result of these problems, there were more valid ratings for the first part of the survey than for the second, which contains relatively many Iambic sentences (five out of 11 vs. four out of 16). Consequently, the data set as a whole was slightly unbalanced. Furthermore, the number of participants (SS) whose two parts of the survey could reliably be combined, was rather small, given that in this experiment there are five timing conditions to be compared. A larger data set (LS), consisting of 68 trials for each part of the survey, was also balanced, but had to be treated as a data set with 102 participants: 34 who rated all stimuli (i.e., SS), plus 34 who rated the first 18 sentences, and 34 who have rated the last 11, although the last two groups consist of largely the same persons. This, of course, weakens the statistical power of the analysis of LS.

Despite this, regressions on the factors showed significant effects of timing within LS. There were also significant effects of timing within All, but, as might have been expected, not within SS (see Table 5). However, the mean factor scores per factor per song version show the same patterns across samples (see Figure 4), and some of them even seem to be more salient within SS than within LS and All (e.g., Rightness early, and both Urgency and Rightness very late). Possibly, the lack of significance of these patterns within SS compared with LS and All is due to sample size. Even the fact that Iambic sentences are underrepresented in the complete data set hardly affects the structure of the models (see Table 6). *F* for the various variables within the model, including timing, is quite similar across samples (see, e.g., Table 6 for the details for Models B for Urgency, Upsetness, and Rightness within LS and All). Moreover, *F* for timing is quite similar in Models A compared to Models B. Therefore, the effect of timing on the singer's perceived tone of voice seems to be largely independent of sentence or melody.

So, notwithstanding the flaws in the data set, the results of Experiment 2 show, as expected, a stronger effect of timing than the results of Experiment 1 and the pilot experiment. Again, on-beat phrasing turned out to support Rightness ($p < .01$), but this time Upsetness and Urgency also showed significant effects ($p < .05$), which is in line with the hypothesis that off-beat phrasing would create some emotional subtext to the words of the sentence. The results indicate that off-beat versions in general and very late ones in particular are more often associated with Upsetness than on-beat versions, and early and very early ones with Urgency. Post hoc pairwise comparisons between the categories of timing revealed that Urgency is higher for very early sentence versions compared to very late ones, (which is remarkable, because the means for early versions are higher than those for very early ones), that on-beat versions are rated as less upset than very late ones, and that very late versions are significantly less right than on-beat and late ones.

Pleasantness did not show a significant effect, which could be due to the fact that too few items in the questionnaire were as yet related to it. It is therefore interesting to see that there is a clear tendency for early and very early sentence versions to be perceived as unpleasant, whereas on-beat, late, and very late ones tend to be perceived as relatively pleasant.

As in the previous experiments, models with aspects of melody as fixed effects are stronger than models with a random intercept for melody (see Table 5). Yet, notably, this does not negatively affect, the results for the effect of timing. An exploratory investigation of the effects of tempo, measure, metre, sentence type, and nonclosure largely, yielded results which resemble those of Experiment 1. They will be further analysed in the General Discussion.

GoldMSI MT and Writing experience could only be used as a covariate within SS. The results indicate that experienced writers tend to rate the stimuli as more Upset and less Right than people without writing experience, and that musicians tend to rate the stimuli as both more pleasant and right. These results are consistent with neither those of Experiment 1, nor those of the pilot experiment.

Figure 3

Examples of the Four New Sentence Versions, Starting From the Top: Early, Late, Very Early, and Very Late

Table 4
Factor Analysis of Affect Ratings, Experiment 2

Item	Factor			
	Pleasant	Urgent	Upset	Right
All sounds insecure.		.108	**.664**	.364
The voice sounds upset.	.234	.258	**.717**	
The singer sounds restless.	.311	**.480**	**.591**	.153
The singer sounds hesitant.			**.560**	.304
The singer sounds sincere.	−.137		−.156	**−.766**
The singer sounds convincing.		.178	−.389	**−.675**
Lyrics and music are a good match.	−.301	−.100	−.160	**−.582**
All sounds stable.	−.340	−.146	−.410	**−.526**
The voice sounds nagging.	.259	**.804**	.140	
The lyrics sound compelling.	.340	**.680**		−.173
The singer sounds greedy.		**.565**	.302	
All sounds hasty.		**.529**	.186	.156
All sounds cheerful.	**−.769**		−.395	−.328
The lyrics sound friendly.	**−.736**	−.318	−.181	−.359
The singer sounds indifferent.	−.109	.164	.246	.331
Initial eigenvalue	3.99	2.43	1.59	1.30
Percentage of variance predicted	26.61	16.20	1.11	0.79
Rotated sum of squared loadings	1.72	2.22	2.38	2.31

Note. Extraction method: Principal axis factoring. Rotation method: Oblimin with Kaiser Normalization. Kaiser-Meyer-Olkin Statistic (KMO) = .805; determinant = .008; Measures of Sampling Adequacy (MSA) for all variables $>.7$.

General Discussion

Following on from an earlier pilot experiment, two experiments have been conducted to test Pattison's (2015) assumption that off-beat phrasing negatively affects musical stability, and that subsequently the singer's tone of voice will be perceived as less sincere and convincing and the message as less straightforward. Both experiments have indicated that musical stability is indeed associated with affective qualities such as sincerity, convincingness, and naturalness, resulting in the emergence of a factor representing both aesthetic and moral Rightness, and which may be related to Aaftink's aesthetic factor "Just rightness." Another factor, called Pleasantness, seems to combine stability with friendliness and cheerfulness, although in Experiment 2 the connection with stability is less convincing.

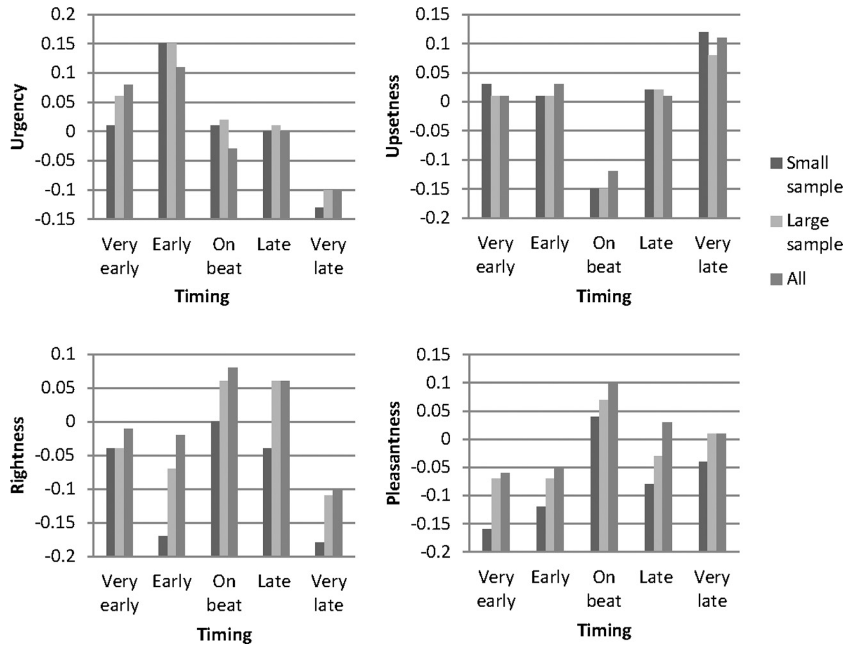
In line with Pattison (2015), on-beat phrasing has turned out to support Rightness, whereas off-beat phrasing causes the singer's tone of voice to be interpreted as being unbalanced in either a more pressing or a more tentative way. Interpretations of these kinds where represented by the factors Urgency and Upsetness, respectively. These factors are reminiscent of the traditional categories of interactive expression, that is, antagonistic and affinitive behavior (Frijda, 1986; Huron, 2015) but do not coincide with them completely, as Pleasantness represents affinitive behavior as well, and Rightness represents aspects of both.

Table 5
Akaike's Information Criterion (AIC) for Various Models Per Sample, Per Factor, and F for Timing in Model A, If Significant

	SS	LS	All	SS	LS	All
	Urgency			*Upsetness*		
Intercept	2,061.72	4,225.53	5,072.96	2,224.42	4,322.79	5,126.75
Model A	2,065.44	4,224.76	5,071.09	2,227.62	4,320.39	5,124.55
F_{timingA}	n.s	2.23^{+}	2.48*	n.s.	2.66*	2.60*
Model A1		4,221.47	5,066.98		4,314.25	5,119.75
Model B	2,050.83	4,209.11	5,055.10	2,210.34[a]	4,308.10	5,112.80
	Rightness			*Pleasantness*		
Intercept	2,345.78	4,639.41	5,465.61	2,069.20	4,060.02	4,802.01
Model A	2,346.69	4,632.84	5,460.49	2,071.56	4,066.24	4,808.64
F_{timingA}	n.s.	3.49**	3.13*	n.s.	n.s.	n.s.
Model B	2,332.98[a]	4,615.16	5,442.29	2,056.98[a,b]	4,055.49	4,798.17

Note. SS = small sample; LS = large sample; All = all participants; n.s. = not significant. Intercept = Intercept only (participant, sentence, melody); Model A = Intercept (with melody) and timing; Model A1 = Intercept (with melody), timing, and sentence type; Model B = Intercept (without melody), timing, and various aspects of sentence type and melody (see notes [a] and [b] and Table 6).
[a] SS Model B for Upsetness including Writing experience (t = 4.16*); for Rightness including GoldMSI MT (Goldsmith Musical Sophistication Index, Musical-Training subscale) (t = 2.26*) and Writing experience ($t = -1.96^{+}$); for Pleasantness including GoldMSI MT (t = 6.10*). [b] Factors predicting pleasantness are meter and nonclosure; a model with tempo and nonclosure is weaker.
* $p < .05$. ** $p < .01$. $^{+}$ $p < .06$.

Figure 4
Effects of Timing



Note. Factor scores per sentence version per sample. SS = small sample, LS = large sample, All = all participants.

Admittedly, the effect of timing on Urgency and Upsetness was only significant when the last stressed syllable in a sung sentence was phrased off beat. There are several explanations for this: in these cases the listener's attention could not shift toward the rhythm of the language (Gordon et al., 2011); the beat was established more clearly before the deviance ocurred, and the off-beat note onset is not only deviant in relation to the beat, but also to the rest of the sentence. This raises the question whether the effect of timing would be different if either the first or the second stressed syllable would be timed off beat, or if the sentences would not have had the same length. Van Heuven and Van Zanten (2005) have shown that, at least, local accelerations and slowdowns within statements can change them into questions.

As I have mentioned before, there is a small chance that on-beat phrasings are rated as more "right" because most of the on-beat stimuli were the original recordings, whereas most of the off-beat stimuli were digitally altered. However, this effect would have been the same for all off-beat stimuli. It cannot explain the differences between early and late phrasings, or between the effect of the stimuli in Experiment 1 and those in Experiment 2. Therefore, I think it is reasonable to conclude that the results are largely due to timing.

These findings are not only in line with Pattison, Gordon, and DAT (Jones, 1976) but also with the musical foregrounding hypothesis (Schotanus, 2015, 2020b), as they show that timing can be used to further a specific interpretation of a song's lyrics. As expected, the "wrongness" of an off-beat occurrence of a major linguistic event seems to be attributed to the singer's intentions or emotional state. This contributes to a growing body of evidence that listeners expect linguistic events matched to unexpected musical events to be deviant as well. For example, nonword recognition is faster when matched with unexpected harmonies (Curtis & Bharucha, 2003), and ambiguous words sung to out-of-key notes are less often interpreted literally (Schotanus, 2018).

Depending on the means, early and very late versions elicit more deviant ratings than very early and late versions respectively, which is in line with the predictions concerning loud-rest strength. Remarkably, however, the deviant ratings for early versions were not significant, whereas the less deviant Urgency ratings for very early versions were. Given the fact that haste-related Urgency ratings were higher for early versions, whereas hesitation-related Upsetness ratings were higher for late ones, the suggestion is that the violation of other predictions interacts with this. At least, it seems to be clear that "late notes" in these stimuli were perceived as syncopated in relation to the main beat, not to the third one.

Remarkably, notwithstanding the notions "hasty" in Urgency, and "hesitant" in Upsetness, Upsetness ratings are also relatively high for early and very early versions, whereas Urgency ratings are relatively high in late ones. These results may reflect the ambiguousness of speech rate as an ethological signal or cue: On the one hand slow speech is associated with dominance (Tusing & Dillard, 2000), but on the other it is also associated with fear and sadness (Siegman & Boyle, 1993). Speech rate probably interacts with other features such as loudness and pitch.

The additional explorative investigations of the effect of various properties of sentences and melodies show remarkably consistent trends across experiments. Although these results should be interpreted with care, because the ratings were not properly counterbalanced, they are often salient and consistent enough to mention and to be discussed in relation to existing theories. It may invite further research with more balanced stimuli.

Table 6
The Effect of Timing and Various Aspects of Sentence Type and Melody on Urgency, Upsetness, and Rightness

	Urgency		Upsetness		Rightness	
	Model B LS	Model B All	Model B LS	Model B All	Model B LS	Model B All
			F			
Intercept	2.03	2.85	0.28	0.19	0.68	1.55
Timing	2.31 (p = .055)	2.58*	2.69*	2.60*	3.50**	3.09*
Measure	6.83*	6.58*	5.29*	5.41*	22.52***	21.66***
Meter	9.97**	10.44**			18.34***	14.71**
Tempo	16.82***	17.52***	27.72***	28.66***		
Sentence type	3.23*	3.56*	4.53*	4.18*	4.78**	4.42*
			t			
Very early	2.62**	2.87**	−1.40	−1.61	1.26	1.55
Early	2.48*	2.39*	−0.13	−0.62	0.56	1.02
On beat	1.04	1.03	−2.78**	−2.90**	2.87**	3.05**
Late	1.46	1.19	−0.37	−0.55	2.93**	2.71**
Very late	r.c.	r.c.	r.c.	r.c.	r.c.	r.c.
4 / 4 time	−2.61*	−2.57*	2.30*	2.33*	4.75***	4.65***
3 / 4 time	r.c.	r.c.	r.c.	r.c.	r.c.	r.c.
Trochee	3.16**	3.23**			−4.28**	−3.84**
Iamb	r.c.	r.c.			r.c.	r.c.
Slow	−4.01***	−4.02***	5.96***	6.01***		
Moderate	−1.28	−1.20	0.28	0.20		
Fast	r.c.	r.c.	r.c.	r.c.		
Imperative	1.11	1.42	0.21	0.18	−1.54	−1.55
Question	−1.23	−1.04	3.06**	2.92**	−1.66	−1.88
Ellipsis	−1.78	−1.75	−0.1	−0.12	−3.77**	−3.60**
Statement	r.c.	r.c.	r.c.	r.c.	r.c.	r.c.

Note. *F* for variables, and *t* for their categories in models with a significant effect of timing. LS = large sample; All = all participants; r.c. = reference category.
* $p < .05$. ** $p < .01$. *** $p < .001$.

First, the use of either iambs or pick-up notes seems to support Rightness ($p_{\text{both experiments [b.e]}} < .01$) and Pleasantness ($p_{\text{b.e.}} < .001$), whereas trochees seemed to support Urgency ($p_{\text{b.e.}} < .01$). It is likely that the pick-up notes help predict the occurrence of stressed syllables (which are the major linguistic events), which would be pleasant and makes the timing feel right. This would also provide evidence for Hansen et al.'s (2018) theories on the effect of dotted rhythms. That trochees would support Urgency may be due to the structure of the Dutch language. According to Bronzwaer (1993), trochees tend to sound like a drone in Dutch, because Dutch words tend to be trochaic as well.

Second, a relatively high tempo seems to support Urgency ($p_{\text{b.e.}} < .001$) while a relatively low tempo seems to support Upsetness ($p_{\text{b.e.}} < .001$). This would be consistent with the notions "hasty" and "holding back" within these factors, and with Siegman & Boyle (1993). However, it would not be consistent with Tusing & Dillard (2000).

Third, the connection that seems to exist between Rightness and a four-four time measure ($p_{\text{b.e.}} < .001$) is in line with both the balanced rhythm and a strong preference for binary structures in Western music (Condit-Schultz, 2020; Tan et al., 2019; Temperley, 2001, p. 39). However, the connections that seem to be visible between Urgency and a three-four time beat on the one hand ($p_{e1} < .01$; $p_{e2} < .05$), and Upsetness and a four-four time beat on the other ($p_{e2} < .05$), are puzzling. It is possible that the sense of a rotating sound, inherent to triplets (Hansen & Huron, 2019), plays a part in this.

Fourth, the surprisingly small effect of sentence type ($p_{\text{e1Rightness}} < .05$; $p_{\text{e2all variables}} < .05$) can be explained by the fact that in Dutch, statements can turn into questions or commands, dependent on the prosody (see, e.g., the declarative questions in Van Heuven & Van Zanten, 2005), and similarly questions can turn into commands and commands into requests. Furthermore, the effect of each sentence's text is specifically integrated into the random effect of the sentence.

Finally, musical training, writing experience, and nonclosure did not show consistent significant effects across studies. In the case of nonclosure, this may partly be due to the fact that closure-related qualia are not dependent on closure strength but on scale-degree qualia (Arthur, 2018). Furthermore, in the current stimuli melody, the accompaniment can diminish a possible sense of nonclosure in the voice part, which nevertheless may affect the interpretation of the singer's tone of voice.

Due to several other limitations of the current study, future research could also reckon with pitch height and adapt the stimuli in several other ways. One could, for example, investigate the effect of varying sentence length and the part of the sentence to be phrased off beat. Furthermore, one could match off-beat phrases more consciously to so-called affect-carrying words (Sun & Cuthbert, 2018) or investigate correlations between early and late phrases and categories of those words.

Future research should also check for dissonances occurring when the melody is shifted in relation to the accompaniment. Therefore, one should look for ecologically valid, varied, and attractive combinations of melody and accompaniment that will not cause dissonances (presumably a simple vamp, or a djembe beat) when the melody is shifted in relation to the accompaniment. In addition, it would be interesting to create a set of stimuli in which off-beat phrases are actually sung off-beat. This will probably strengthen the effect, as singers may change their prosody if they are singing off-beat intentionally. Another option would be to create a set of stimuli in which for each melody one sentence is sung on-beat, one early and one late in the original recording, to control for effects of both prosody and digital modification.

Finally, future brain-imaging research may investigate the occurrence of brainpotentials such as an MMN (mismatch negativity) on the potential loud rest in off-beat versions, of a, so called, N400 (a negativity indicating meaningfulness) in reaction to words phrased off-beat, or of any brain potential that could be associated with a sustained loud rest in very late versions.

Given the number of regression analyses presented above, one could ask whether there are no problems with multiple comparison. However, I do not think so. In Experiment 1, there has been one factor analysis, after which four factors were analyzed in the same way. There was just one hypothesis-based target regression per factor concerning the effect of timing, the validity of which was tested afterward by comparing it with other models, which is fair. What is more, if one would want to correct the *p* factor for the main effect of timing on Rightness allows by multiplying it by 4, the effect would still be significant. It is true that I have also reported the stronger alternative models which allowed me to discuss the effect of several aspects of melody. However, this was done only in an exploratory way. Moreover, only by reporting the results for the stronger models, I was able to show that the effect of timing on Rightness holds even within the context of the alternative model. So, reporting these extra regression results does not weaken the conclusions but strengthens them. Something similar holds for Experiment 2, except that in this experiment, due to problems with the data set, separate analyses on subsets of it were required to check the validity of the analyses on the whole data set. As the results for LS do not differ substantially from the results for All, these analyses strengthen the results for All, rather than weakening them.

Conclusions

In summary, in two online listening experiments, timing is shown to affect the perception of the singer's tone of voice. If stressed syllables occur on beat, the singer is perceived as being relatively "right" (a combination of sincerity, naturalness, and convincingness, among other things), if they precede the main beat, the singer's tone of voice is perceived as relatively Urgent, and if they succeed the main beat, as rather Upset. Several aspects of melody turned out to affect these factors as well. The results can be related to various theories, shed light on aspects of syncopation and emotions in music that have scarcely been studied yet, and indicate that the extent to which a singer is perceived to be "authentic," which is very important in the music industry, can be modified by using specific musical features.

References

Aaftink, C. (2014). *Kaleidoscope: A phenomenological-empirical study of beauty* [Doctoral dissertation]. University of Alberta.

Arthur, C. (2018). A perceptual study of scale-degree qualia in context. *Music Perception*, *35*(3), 295–314. https://doi.org/10.1525/mp.2018.35.3.295

Auslander, P. (1999). *Liveness: Performance in a mediatized culture*. Routledge.

Bouwer, F. L., & Honing, H. (2015). Temporal attending and prediction influence the perception of metrical rhythm: Evidence from reaction times and ERPs. *Frontiers in Psychology*, *6*, Article 1094. https://doi.org/10.3389/fpsyg.2015.01094

Bouwer, F., Schotanus, Y. P., Sadakata, M., Müllensiefen, D. & Schaefer, R. (2021). *Measuring musical sophistication in the low countries; validation of a Gold MSI translation in Dutch*. Manuscript submitted for publication.

Brackett, D. (1995). *Interpreting popular music*. Cambridge University Press.

Bronzwaer, W. (1993). *Lessen in lyriek: Nieuwe Nederlandse poetica* [Lessons on the lyric: New Dutch poetics]. SUN.

Burns, L. (2000). Analytic methodologies for rock music: harmony and voice leading strategies in Tori Amos's 'Crucify. In W. Everett (Ed.), *Expression in pop-rock music: A collection of critical and analytical essays* (pp. 213–246). Garland.

Cespedes-Guevara, J., & Eerola, T. (2018). Music communicates affect, not basic emotions: A constructionist account of attribution of emotional meanings to music. *Frontiers in Psychology*, *9*, Article 215. https://doi.org/10.3389/fpsyg.2018.00215

Condit-Schultz, N. (2020). Expanding and contracting definitions of syncopation: Commentary on Temperley 2019. *Empirical Musicology Review*, *14*(1-2), Article 81086. https://doi.org/10.18061/emr.v14i1-2.7098

Curtis, M. E., & Bharucha, J. J. (2003). Tonal violations interact with lexical processing: Evidence from cross-modal Priming [Paper presentation]. *Annual conference of the Society for Music Perception and Cognition, Las Vegas, Nevada.*

Eckstein, L. (2010). *Reading song lyrics*. Rodopi. https://doi.org/10.1163/9789042030367

Findeisen, F. (2017). *How Ed Sheeran writes a melody: The artists series S1E3* [YouTube lecture]. Holistic songwriting. https://www.youtube.com/watch?v=LWQVztiJHfs

Frijda, N. (1986). *The emotions*. Cambridge University Press.

Frith, S. (2007). 'The magic that can set you free': The ideology of folk and the myth of the rock community. In S. Frith (Ed.), *Taking popular music seriously: Selected essays* (pp. 31–40). Ashgate. (Re-printed from: Popular music 1, 159–168, by R. Middleton & D. Horn, Eds., 1981, Cambridge University Press).

Frith, S. (2007). What is bad music. In S. Frith (Ed.), *Taking popular music seriously: Selected essays* (pp. 313–334). Ashgate. (Reprinted from: Bad music: The music we love to hate, pp. 15–36, by C. J. Washburne & M. Derno, Eds., 1987). https://doi.org/10.4324/9780203309049_chapter_1

Gabrielsson, A., & Juslin, P. N. (1996). Emotional expression in music performance: Between the performer's intention and the listener's experience. *Psychology of Music*, *24*(1), 68–91. https://doi.org/10.1177/0305735696241007

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*, *2*, Article 352. https://doi.org/10.3389/fpsyg.2011.00352

Hansen, N. C., & Huron, D. (2019). Twirling triplets: The qualia of rotation and musical rhythm. *Musicae Scientiae*, *2*, 205920431881224. Advance online publication. https://doi.org/10.1177/2059204318812243

Hansen, N. C., Shea, N. J., & Huron, D. (2018). Do dotted rhythms increase performance precision: Why marches have dotted rhythms. In R. Parncutt & S. Sattmann (Eds.), *Proceedings of ICMPC15/ESCOM10* (pp. 184–189). Centre for Systematic Musicology University of Graz.

Huron, D. (2015). Cues and signals: An ethological approach to music-related emotion. *Signata*, *6*(6), 331–351. https://doi.org/10.4000/signata.1115

Huron, D. (2016). *Voice leading: The science behind a musical art*. Cambridge (MA): MIT.

Huron, D., Kinney, D., & Precoda, K. (2006). Influence of pitch height on the perception of submissiveness and threat in musical passages. *Empirical Musicology Review*, *1*(3), 170–177. https://doi.org/10.18061/1811/24068

Huron, D., & Royal, M. (1996). What is melodic accent? Converging evidence from musical practice. *Music Perception*, *13*(4), 489–516. https://doi.org/10.2307/40285700

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–355. https://doi.org/10.1037/0033-295X.83.5.323

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770–814. https://doi.org/10.1037/1528-3542.2.1.23

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Reviews*, *8*(2), 89–105. https://doi.org/10.1016/j.plrev.2011.04.004

Koops, H. V., Volk, A., & De Haas, W. B. (2015). Corpus-based rhythmic pattern analysis of ragtime syncopation. *Proceedings of the 16th International Society for Music Information Retrieval Conference*, pp. 483–489. ISMIR.

Ladinig, O., Honing, H., Háágden, G., & Winkler, I. (2009). Probing attentive and pre-attentive emergent meter in adults and adult listeners with no extensive music training. *Music Perception*, *26*(4), 377–386. https://doi.org/10.1525/mp.2009.26.4.377

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time varying events. *Psychological Review*, *106*(1), 119–159. https://doi.org/10.1037/0033-295X.106.1.119

London, J. M. (1993). Loud rests and other strange metric phenomena (or, meter as heard). *Music Theory online*. https://mtosmt.org/issues/mto.93.0.2/mto.93.0.2.london.php

Longuet-Higgins, H. C., & Lee, C. W. (1984). The rhythmic interpretation of monophonic music. *Music Perception*, *1*(4), 424–441. https://doi.org/10.2307/40285271

Menon, V., & Levitin, D. J. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *NeuroImage*, *28*(1), 175–184. https://doi.org/10.1016/j.neuroimage.2005.05.053

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2), Article e89642. https://doi.org/10.1371/journal.pone.0089642

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, *41*(1), 1–16. https://doi.org/10.1159/000261706

Pattison, P. (2015). *Lesson 46: Phrasing. Songwriting: writing the lyrics*. Coursera.

Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica*, *62*(1), 1–13. https://doi.org/10.1159/000087222

Scherer, K., Trznadel, S., Fantini, B., & Sundberg, J. (2017). Recognizing emotions in the singing voice. *Psychomusicology: Music, Mind, and Brain*, *27*(4), 244–255. https://doi.org/10.1037/pmu0000193

Schotanus, Y. P. (2015). The musical foregrounding hypothesis: How music influences the perception of sung language. In J. Ginsborg, A. Lamont, M. Philips, & S. Bramley (Eds.), *Proceedings of the Ninth Triennial Conference of the European Society for the Cognitive Sciences of Music, 17–22 August 2015*, Manchester, U.K.

Schotanus, Y. P. (2018). Off-beat phrasing and the interpretation of the singer's tone of voice. In R. Parncut & S. Sattmann (Eds.), *Proceedings of ICMPC15/ESCOM10* (pp. 401–406). Centre for Systematic Musicology, University of Graz.

Schotanus, Y. P. (2019). "Supplementary materials PhD project Singing as a figure of speech and related publications", *DataverseNL, V1*. https://hdl.handle.net/10411/XRMMFW

Schotanus, Y. P. (2020a). Singing and accompaniment support the processing of song lyrics and change the lyrics' meaning. *Empirical Musicology Review*, *15*(1-2), 18–55. https://doi.org/10.18061/emr.v15i1-2.6863

Schotanus, Y. P. (2020b). *Singing as a figure of speech, music as punctuation: A study into music as a means to support the processing of sung language* [Doctoral dissertation]. Utrecht University. https://doi.org/10.33540/249

Schubert, E. (2013). Emotion felt by the listener and expressed by the music: Literature review and theoretical perspectives. *Frontiers in Psychology*, *4*, Article 837. https://doi.org/10.3389/fpsyg.2013.00837

Shanahan, D., & Huron, D. (2014). Heroes and villains: The relationship between pitch tessitura and sociability of operatic characters. *Empirical Musicology Review*, *9*(2), 46–59. https://doi.org/10.18061/emr.v9i2.4441

Siegman, A. W., & Boyle, S. (1993). Voices of fear and anxiety and sadness and depression: The effects of speech rate and loudness on fear and anxiety and sadness and depression. *Journal of Abnormal Psychology*, *102*(3), 430–437. https://doi.org/10.1037/0021-843X.102.3.430

Sun, S. H., & Cuthbert, M. S. (2018). Emotion, painting: Lyric, affect, and musical relationships in a large lead-sheet corpus. *Empirical Musicology Review*, *12*(3-4), 327–348. https://doi.org/10.18061/emr.v12i3-4.5889

Tan, I., Lustig, E., & Temperley, D. (2019). Anticipatory syncopation in rock: A corpus study. *Music Perception*, *36*(4), 353–370. https://doi.org/10.1525/mp.2019.36.4.353

Temperley, D. (2001). *The cognition of basic musical structures*. MIT.

Temperley, D. (2009). A unified probabilistic model of polyphonic music analysis. *Journal of New Music Research*, *38*(1), 3–18. https://doi.org/10.1080/09298210902928495

Temperley, D. (2019). Second-position syncopation in European and American vocal music. *Empirical Musicology Review*, *14*(1–2), 66–80. https://doi.org/10.18061/emr.v14i1-2.6986

Tusing, K. J., & Dillard, J. P. (2000). The sounds of dominance: Vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, *26*(1), 148–171. https://doi.org/10.1111/j.1468-2958.2000.tb00754.x

van Heuven, V. J., & Van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication*, *47*(1-2), 87–99. https://doi.org/10.1016/j.specom.2005.05.010

Warrenburg, L. (2020). Comparing musical and psychological emotion theories. *Psychomusicology: Music, Mind, and Brain*, *30*(1), 1–19. https://doi.org/10.1037/pmu0000247

Witek, M. A. G., Clarke, E. F., Wallentin, M., Kringelbach, M. L., & Vuust, P. (2014). Syncopation, body-movement and pleasure in groove music. *PLoS ONE*, *9*(4), Article e94446. https://doi.org/10.1371/journal.pone.0094446

Witek, M. A. G., Clarke, E. F., Kringelbach, M. L., & Vuust, P. (2014). Effects of polyphonic context, instrumentation, and metrical location on syncopation in music. *Music Perception*, *32*(2), 201–217. https://doi.org/10.1525/mp.2014.32.2.201

Appendix A

Table A1
Distribution of Sentences, Melodies, and Stimulus Versions Over Stimulus Sets

Melody	Experiment 1			Experiment 2					Sentence - Translation
	Set 1	Set 2	Set 3	Set 1	Set 2	Set 3	Set4	Set 5	
100	e	o	l	ve	vl	o	e	l	Liefste, liefste, *blijf*. – Darlin', darlin', *stay*.
140 (=130)	l	e	o	vl	o	e	l	ve	Schieten, schieten, *schiet*! – Shoot, shoot, *shoot*!
165(=150)	o	l	e	l	e	ve	[l]	o	Kijk eens, jongens, *goud*. – Look here, guys, *gold*.
180 (=170)	e	o	l	o	l	vl	ve	e	Blijf je, blijf je, *toe*? – You stay, you stay, *please*?
205 (=190)	l	e	o	e	ve	l	o	vl	Zeg het, zeg het, *nu*. – Say it, say it, *now*.
210	o	l	e	ve	vl	o	e	l	Deur groen, luiken *geel*. – Door green, shutters *yellow*.
240 (=220)	e	o	l	vl	o	e	l	ve	Verliefd, ik ben, *verliefd*. – In love, I am, *in love*.
250	l	e	o	l	e	ve	vl	o	Schenk in, er is *genoeg*. – Pour, there is *enough*.
290 (=280)	o	l	e	o	l	vl	ve	e	Voor mij één bol *pistache* – For me, one scoop of *pistachio*.
130	e	o	l	e	ve	l	o	vl	Lopen, lopen, *snel*. – Run, run, *fast*.
120 (=100)	l	e	o	ve	vl	o	e	l	Strakjes zie ik *jou*. – Soon, I will see *you*.
160 (=150)	o	l	e	vl	o	e	l	ve	Niemand komt hier *langs*. – No one is passing *by*.
185 (=170)	e	o	l	l	e	ve	vl	o	Papa, papa, *kijk*. – Daddy, Daddy, *look*.
215 (=210)	l	e	o	o	l	vl	ve	e	Ik wil morgen *gaan*. – Tomorrow, I want to *go*.
190	o	l	e	e	ve	l	o	vl	Stemmen, stemmen, *hoor*. – Voices, voices, *hear*.
230 (=220)	e	o	l	ve	vl	o	e	l	Wat nu, heb jij *een plan*? – What next. Do you have *a plan?*
280	l	e	o	vl	o	e	l	ve	Mevrouw, heeft het *gesmaakt*. – Miss, did it *taste well*?
270 (=250)	o	l	e	l	e	ve	vl	o	Geniet, mijnheer, *geniet*. – Enjoy yourself, sir, *enjoy*.
150	e	o	l	o	l	vl	ve	e	Zeg jij dat nou *echt*? – Is that what you say, *really*?
110 (=100)	l	e	o	e	ve	l	o	vl	Morgen sla ik *toe*. – Tomorrow, I will *strike*.
145 (=130)	o	l	e	ve	vl	o	e	l	Toe dan, toe dan, *spring*. – Come on, come on, *jump*.
200 (=190)	e	o	l	[l]	o	e	l	ve	Waar is toch dat *boek*. – Where is that *book*.
217 (=210)	l	e	o	l	e	ve	vl	o	Veerweg, 40, *Velp* – Veerweg, 40, *Velp*.
170	o	l	e	o	l	vl	ve	e	Hector, Hector, *hier*. – Hector, Hector, *here*!
300 (=280)	e	o	l	e	ve	l	o	vl	Vandaag doen we *het huis*. – Today we will clean *the house*.
220	l	e	o	ve	vl	o	e	l	Kom op, vertel, *vertel*. – Come on, tell us, *tell*.
260 (=250)	o	l	e	vl	o	e	l	ve	Tast toe, dit is *iets nieuws*. – Get some, it is *something new*.

Note. e = early; o = on beat; l = late; ve = very early; vl = very late. The manipulated words and their equivalents are in italics. In case of an iamb, two words have sometimes been manipulated.

(*Appendices continue*)

Appendix B

Figure B1

The Eight Melodies Not Depicted Within the Main Body of the Article, for the Ninth Melody See Page 110.

Received September 1, 2020
Revision received May 26, 2021
Accepted June 2, 2021 ■