In Parncutt, R. & Sattmann, S. (Eds). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz. 401-406.

Off-Beat Phrasing and the Interpretation of the Singer's Tone of Voice

Yke Schotanus[1]

[1] Institute for Cultural Inquiry (ICON), Utrecht University, The Netherlands

[1]schotschrift.teksten@planet.nl

Abstract

Both music and lyrics are thought to affect the emotional meaning of a song, but to date it is not exactly clear how and to what extent music does so. Possibly, timing is an important factor. Both singers and composers often create off-beat onsets of important linguistic events, such as the first stressed syllable in a phrase (henceforward: phrase onset). However, off-beat events are more difficult to process, which is hypothesized to cause a foregrounding effect, which would affect the interpretation of the singer's state of mind, his or her intentions, and the meaning of the words. An online listening experiment was created to test this hypothesis. Thirty participants listened to 27 piano-accompanied sung sentences, consisting of five or six syllables, some of them statements, some of them questions, imperatives, or incomplete sentences. In nine of them the phrase-onset was on-beat, in 9 it was early and in the remaining 9 it was late. After each sentence participants rated 10 items concerning the way words, music and singer are perceived. Three factors emerged from a factor analysis on these ratings. Regressions on these factors show that they are hardly affected by timing. Surprisingly, also sentence type did just marginally affect one factor. This indicates that music is more important in communicating aspects of meaning such as sincerity, self-security or compellingness.

Introduction

Music is thought to affect the meaning of a song, or at least the emotional meaning. Ali and Peynircioğly (2006) even conclude that music has more impact than lyrics when it comes to perceived happiness. However, to date it is not exactly clear in which ways music does so, although 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 (for example: Gabrielsson & Juslin, 1994; Scherer et al., 2017). To my knowledge there are just a few studies regarding the way music influences perceived affect, for example perceived sincerity, sociability or submissiveness (Huron, Kinney & Precoda, 2006; and Shanahan & Huron, 2014). Nevertheless, perceived affect, is very important feature in popular music especially when it comes to perceived authenticity (Bracket, 1995). According to Pattison (2015) off-beat phrasing (starting the first stressed syllable of a linguistic phrase on a weak beat) is a useful technique to manipulate affect. And indeed, both singers and composers often create such phrase onsets (Temperley, 2001). Pattison argues that off-beat phrases sound less stable, which makes the listener feel that there might be some subtext to the lyrics, for example because the singer is upset.

These presumptions are in line with several theories. In the first place Dynamic Attending Theory (DAT) (Jones, 1976; Large & Jones, 1999), assumes that our attention oscillates in accordance with a given speech rhythm or musical rhythm. Thus, in Western listeners attention is optimal at strong beats, and in a stress-timed language like Dutch it is optimal at stressed syllables. In line with that, several studies have shown that stressed syllables presented off-beat were more difficult to process than stressed syllables presented on-beat (Gordon, 2011; Quené & Port, 2005). Moreover, off-beat phrasings often cause loud rests, or on-beat silences (London, 1993). Such loud rests are known to cause substantial brain activity (Honing et al., 2009), which might distract from language processing. So, both Dynamic Attending and loud-rest processing might obstruct the processing of the words.

However, there is more to it. In literary research, obstructions of the processing of language through stylistic features is known to enhance the perceived salience of either the message or the wording if and only if the obstruction can be interpreted as meaningful (Miall & Kuiken, 1994; Hakemulder, 2004). In line with that The Musical Foregrounding Hypothesis (Schotanus, 2015) states that obstructions of language processing caused by musical events such as off-beat phrasings are, if possible, interpreted as meaningful prosodic cues. The timing of the stimulus, or the confusing of the listener, might be attributed to the intentions or the emotional state of the singer, who might be perceived to be hasty, compelling, lingering, or uncertain, or something like that. Furthermore, the salience of either words or melody might be increased and positively valued. These assumptions were tested in an online listen experiment, with on-beat phrases, early phrases (one eighth before the down beat), and late phrases (on the second beat). On-beat phrases were hypothesized to sound more convincing. The singer will be perceived as more sincere, less insecure, less emotional, and less compelling. On the other hand, on beat phrases might be more predictable and stable and therefore less interesting, and less emotionally loaded (Menon & Levitin, 2005).

Method

Participants and Procedure

Thirty participants (between 18 and 87 years of age; M 55.4; SD = 17.8) were recruited through social media and websites such as proefbunny.nl. They listened at a self-chosen time and place to 27 sung sentences, preceded and accompanied by a piano accompaniment (total track duration about 12 seconds). After each sentence they rated on a seven point likert scale whether they agreed with the statements that: the singer was sincere, insecure, and sounded compelling; whether the lyrics were emotional, superficial, and sounded natural; whether the fragment sounded loaded and energetic; and whether the melody was interesting, and music and lyrics were a good match. Although the hypotheses mainly concerned the interpretation of the singers tone of voice, there

were just a few questions that directly addressed the singer, in order to mask the aim of the experiment. A factor analysis was planned to unveil the connections between the ratings. As participants had to rate 27 stimuli, the number of questions had to be limited.

Half way through the experiment participants answered a few questions on their musical and literary training. Musical training was measured with a translation in Dutch of the Musical training subscale of the Gold MSI (ranging from 7 to 49, Bouwer et al., in preparation; Müllensiefen et al., 2015). Participants score ranged between 7 and 45 (Mean 24,5; SD 9.5). From a factor analysis on the literary-training items, two factors emerged: literary training and disinterest in wording. Musical training, literary training and disinterest in wording were used as covariates in the regressions. The whole experiment lasted about half an hour, participants who completed the whole questionnaire and left behind their address at the end of it received a 5€ book gift voucher.

Stimuli

27 sentences were sung to 9 melodies. There were 13 directives, 5 questions, 6 statements, and 3 elliptic sentences (for example an address, or: Door red, shutters green. All sentences consisted of 3 metrical feet; 18 of Trochees, 9 of Iambs. Hence, the former consisted of 5 syllables and the latter of 6. The melodies were created by the author to express feelings appropriate to at least one of the sentences sung to it, and to vary in contour, key and harmony, as off-beat phrasing was thought to have a general effect independently of text and music. They also varied in measure: 6 melodies were in three-four time, three in four-four time. The sung melodies were preceded and accompanied by piano music improvised by Christan Grotenbreg, a professional musician. He was asked to create different kinds of accompaniments, whether or not using the harmonies suggested by the author, but always establishing a beat. Hence, all sentences would have a clear and similar rhythmic structure, aligned to a well established beat, but would sound relatively interesting and ecologically valid, as far as possible given the atomic design of the study.

In most cases neither the melody nor the accompaniment ended on a tonic. In order to measure harmonic closure for all of the melodies forty Amazon Mechanical Turk workers who did not speak Dutch and could not address the contents of the sentences were presented an example of each melody and were asked to rate whether the fragment sounded like it was 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 non-closure (see Appendix).

All sentences were sung once, but 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. There were no fillers created, as it is impossible not to time either off beat or on beat. Afterwards, the 81 musical fragments, were distributed over three sets of stimuli, such that each participant heard each sentence once, and each melody once in each version.

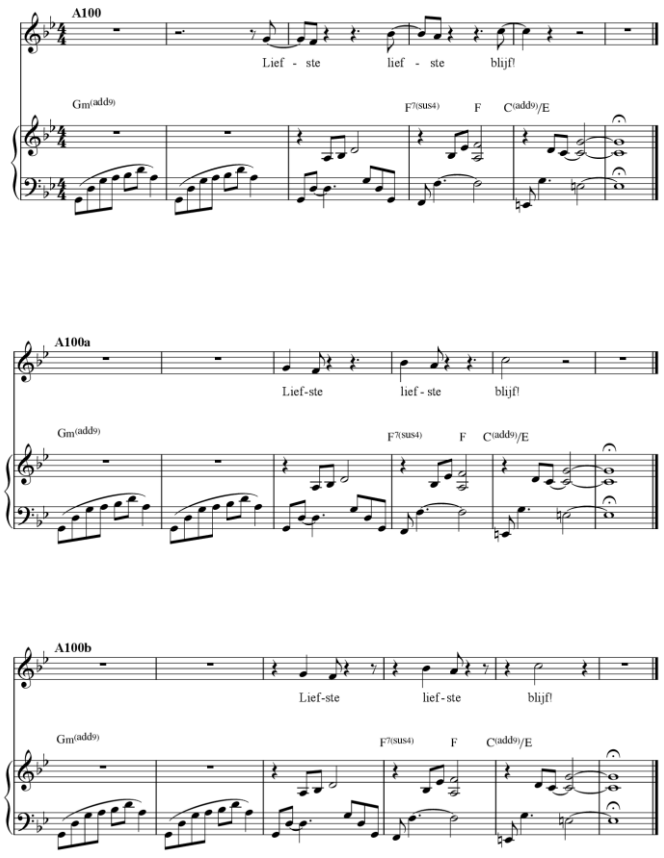

Figure 1. Stimulus example. One sentence 'Liefste, liefste, blijf' [Darlin', Darlin', stay.] In three versions: Early (A100); On beat (A100a) and Late (A100b).

All sentences were sung by the author (a male baritone), 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 confounds 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. After a Principal Axis factor analyses on the ten ratings for each sentence, three factors emerged with an eigenvalue higher than 1 (see Table 1.): Rightness, a combination of naturalness, sincerity, aesthetic valence, and, to a lesser extent, energy, emotionality, and self-security; Upsetness, a combination of insecurity, emotionality and a lack of energy; and Compellingness, a combination of emotional load, and compellingness.

In Parncutt, R. & Sattmann, S. (Eds). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz. 401-406.

Table 1. Factor analysis on the ten ratings per sentence[a].

	Rn[b]	Un[c]	Cn[d]
Initial Eigenvalue	3.51	1.85	1.04
% of variance predicted	35.12	18.47	10.38
Rotation sums square loadings	2.91	1.29	1.75
Factor loadings			
Lyrics sound natural	**.77**	-.19	.37
Singer seems to be sincere	**.72**	-.12	.38
Music and lyrics are good match	**.72**	-.19	.20
Melody is interesting	**.70**	-.19	.21
Singer is insecure	-.24	**.58**	
Fragment sounds energetic	.43	**-.57**	.22
Lyrics are emotional	.46	**.52**	.49
Fragment sounds loaded	.40	.48	**.77**
Singer sounds compelling	.26		**.67**
Lyrics are superficial	-.34		-.22

[a] KMO .78; $p < .001$; df 45; determinant .04
[b] Rightness
[c] Upsetness
[d] Compellingness

Descriptives and Regressions

Depending on Figure 2. Rightness is rated relatively high and Compellingness relatively low for on-beat versions. Upsetness ratings are relatively high for early versions, but the differences are smaller. Whether any of these results are significant or not, was tested in a series of regression analyses. First an intercept only model was tested on each factor in which random intercepts were estimated for participant, sentence and melody. After that models were created using the following factors and covariates: condition (early, on beat or late), meter (Iamb or Trochee), measure (binary or ternary), sentence type (directive, statement, ellipse, question), tempo (slow, mixed, fast), non-closure, musicianship, writing experience, disinterest in wording and interactions between these factors and covariates. Predictors that did not show a significant effect were left out, except for condition. Alternative regressions with melody (df = 8) as a predictor instead of a random factor are slightly less powerful. These models are similar, except that aspects of melody such as meter, measure, tempo and non-closure were redundant and had to be deleted, from these regressions, just as sentence type.

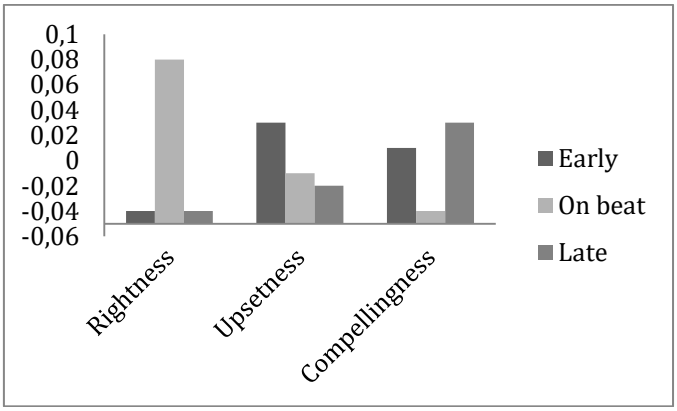

Figure 2. Factor Means per condition. Especially on-beat ratings for Rightness seem deviant. Corresponding SDs Rightness: 1,05; 0.99; 0.91; Upsetness: 1.02; 1.02; 0.96; and Compellingness: 1.0; 1.02; 1.02.

Table 2. Crossed classified linear regression on rightness.

Models	**df**	**-2LL[a]**	**AIC[a]**
Intercept only	5	2013.64	2023.64
Full model	14	1950.72	1978.72
Details		Estimate (SE)	F/Z
Fixed			
Intercept	1		2.43
Condition[b]	2		3.01^{+}
Meter[b]	1		12.56**
Measure[b]	1		10.69*
Cond*Meter[b]	2		1.92
Writingexp.[b]	1		8.22**
Upsetness[b]	1		6.73*
Compellingness[b]	1		4.61*
Random			
Participant		0.19 (0.06)	3.48**
Sentence		0.08 (0.03)	2.40*
Melody		0.01 (0.02)	0.31

[a] Model fit indicators: -2 Log likelihood and Akaikes information criterion
[b] Estimates (SE), if not redundant, for condition: Early: 0.00 (0.11); On beat: 0.26* (0.11); for Meter: Trochee: -0.42* (0.16); for Measure: three-four time: -0.45* (0.14); Late*Trochee: 0.03 (0.14); On beat*Trochee: -0.25 (0.14); Writingexperience -0.25** (0.09); Upsetness 0.10* (0.04), Compellingness 0.07* (0.03).
$^{+}$ $p = 0.05$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Rightness. The effect of timing on rightness does not show a significant main effect unless an interaction with syllable count is included in the model, although even then the effect is only marginally significant ($p = .05$). Furthermore, sentence type, tempo, non-closure, musicianship nor disinterest in wording significantly affected Rightness.

Table 3. Crossed classified linear regression on upsetness.

Models	**df**	**-2LL[a]**	**AIC[a]**
Intercept only		1791. 39	1801.39
Full Model[b]	14	1717.57	1743.57
Details full model		Estimate	F / Z
Fixed			
Condition[c]	2		0.39
Measure[c]	1		11.43**
Tempo[c]	2		92.31***
Non-closure[c]	5		4.75*
Musicianship[c]	1		9.11**
Writing experience[c]	1		8.49**
Disinterest wording[c]	1		6.92**
Random			
Participant		0.06 (0.19)	2.98**
Sentence		0.05 (0.17)	2.73**

[a] Information criteria: -2 Log likelihood, and Akaike's information criterion (-2ll adjusted for model complexity)
[b] Model without random intercept for Melody; model with Melody, but without nonclosure was slightly weaker: df 13, -2ll: 1720.31; AIC: 1746.31
[c] Estimates (Standard Error), if not redundant: Late 0.03 (0.06); On beat -0.02 (0.06); Measure=3: -0.35** (0.10); Slow: 1.55*** (0.13); Mixed tempo 0.35* (0.13); Non-closure: -0.30* (0.14); Musicianship -0.02** (0.05); Writing experience -0.15** (0.05); Disinterest wording 0.08** (0.03).

Upsetness. Timing does not affect Upsetness significantly, nor do sentence type, meter, Rightness and Compellingness. Tempo is by far the main predictor, indicating that slower melodies sound more upset than faster ones. On the other hand, three-four time, and non-closure make the fragments sound more upset. Finally, people who are either trained as writer or as musician tend to give slightly lower upsetness ratings, while people who are not interested in wording give higher ones.

Compellingness. In both Model A (p = .064) and B (see Table 4) the main effect of timing on compellingness ratings approaches significance. However the interaction between timing and musicianship is significant in both models, indicating that musicians tend to rate on-beat sentences as less compelling than late ones. Musicians also tend to rate sentences in general as less compelling. Furthermore, iambs seem to be more compelling than trochees, and quick melodies more compelling than slow ones. Writing experience and measure show no significant effect. Finally, open ended melodies tend to increase compellingness ratings, an effect that approaches significance.

Discussion

The main aim of this study was to test whether off-beat phrasing, as an example of musical instability, affects the emotional meaning of a sung sentence, especially considering the singer's tone of voice. The results indicate that indeed there are differences between early, on-beat and late sentences but they are only marginally significant. However, the results do show that music does affect the interpretation of the singer's tone of voice, and indicate that musical stability might play an important part.

Table 4. Crossed classified linear regression on compellingness.

Models	**df**	**-2LL[a]**	**AIC[a]**	
Intercept only	5	1979.58	1989.58	
Model	14	1910.08	1946.08	
Details Model B		Estimate	F / Z	*p*
Fixed				
Intercept	1		1.25	.271
Condition[b]	2		2.65	.071
Meter[b]	1		10.79	.003
Tempo[b]	2		4.14	.027
Nonclosure[b]	1		3.85	.060
Musicianship[b]	1		4,73	.038
Musician.*cond.[b]	2		4.21	.015
Rightness	1		5.56	.019
Sentence type[b]	3		3.42	.031
Random				
Participant		0.15 (0.04)	3.34	.001
Sentence		0.08 (0.03)	2.95	.003
Melody		redundant		

[a] AIC: 1946.12; variables in model: Melody, Condition, Musicianship, Musicianship*condition, Rightness

[b] Estimates (Standard Error) if not redundant: Late -0.01 (0.18); On beat 0.34^{+} (0.18); Trochee: 0.84** (0.25); Slow: -0.73* (0.33); Mixed tempo 0.21 (0.24); Non-closure: 0.39^{+} (0.20); Musicianship: -0.01 (0.01); Late*Musicianship: 0.00 (0.01); On beat*Musicianship: -0.02* (0.01); Sentence type: Directive: 0.38* (0.17); Statement: 0.27 (0.19); Incomplete -0.26 (0.26); Rightness: -0.08* (0.03).

$^{+}$ $p \leq 0.06$; * $p < 0.05$; ** $p < 0.01$; $p < 0.001$

After each sentence participants rated to what extent they agreed with ten short statements. A factor analysis on these ratings resulted in three factors. The first factor was called rightness, as it seems to indicate that the stimulus was both aesthetically and morally right. Rightness (or 'just rightness') is known as a factor in aesthetics (Aaftink, 2014), but without the moral implication of sincerity. The second factor was called upsetness, as it seems to indicate that the singer was insecure because he was emotional and lacked energy. Finally, the third factor was called compellingness, as compellingness was the main contributor.

It is not easy to connect these factors to the original predictions, but given the part that naturalness and sincerity play in rightness, and the part that insecurity and compellingness play in upsetness and compellingness, one would expect rightness ratings to be higher on beat, and upsetness and compellingness to be higher off beat, which is indeed the pattern shown in Figure 2. However, aesthetic valence was hypothesized to be related to off-beat phrasings which is not in accordance with this pattern. Furthermore, the pattern is not convincingly significant.

The regression on rightness does show a close to significant (p = .050) main effect of condition, but only if an insignificant interaction with meter is involved indicating that trochaic sentences, i.e. melodies without a pick-up note, are rated less right in on-beat versions. Hence, the pick-up note in Iambic sentences might play a substantial part in the effect of condition. One explanation might be that the pick-up note accentuates the beat in the melody and thus the misalignment with the accompaniment. Another explanation might be that in late versions the pick-up notes are dissonant with the accompaniment. This would be an extra dissonant on top of the one often caused by the stressed syllable in early sentences. However, these dissonants do not occur in each sentence, and if they do they occur mostly either in early or in late sentences, while there is no difference in rightness ratings between early and late ones. Furthermore, in popular music people it is normal that melodies are syncopated while the accompaniment is not, and that syncopations are perceived in a different way than in so-called 'classical music' (Temperley, 2001, 239-247; Burns, 2000). Therefore, it is more likely that the main effect of condition is due to Dynamic attending and/or loud rests.

The relatively small effect of condition in all regressions might be due to the fact that the complete melodic line was shifted in relation to the accompaniment. Gordon (2011) found that beat tracking shifts to the rhythm of stressed syllables instead of that of strong beats if stressed syllables consistently occur on weak beats. The effect size could also have been attenuated because the perceived instability might be interpreted and valued in very different ways, dependent on the specific sentence and the melody. Probably, the items to be rated should either be sentence specific, or more clearly aimed at measuring a difference between musical and mental stability, and at clearer factors (for example: convincingness, credibility, sincerity, straightness (i.e. absence of subtext), naturalness, stability, self-security, unrest, upsetness, compellingness, annoyingness, etc.).

Although the regression analyses presented here, were slightly more powerful than regressions with melody as a fixed factor, they have to be interpreted with caution as

neither the aspects of melody nor the categories of sentence type are counterbalanced sufficiently, while melody is. However, in an explorative way, these regressions are more informative than the regression with melody, as they give indications why certain melodies make a singer sound more sincere, more upset, or more compelling. Tempo (or note rate), seems to affect compellingness and upsetness in opposite directions, probably as an indicator of energy. Meter seems to affect compellingness and rightness in opposite directions, possibly because the pick-up note in Iambs enhances predictability and stability. Either the familiarity with four-four time beats, or a general preference for binary structures (Temperley, 2001, 39) might explain why a three-four time decreases both rightness and upsetness ratings. Finally, as dominant-endings are perceived as relatively open it might come as no surprise that non closure increases compellingness, and increases upsetness, although one would expect upsetness not to be associated with complete closure.

Furthermore, the fact that sentence type had to be deleted from all models except one is rather surprising. Admittedly, just as several aspects of melody, sentence type was not properly counterbalanced, but nevertheless the results of this experiment indicate that melody affects the emotional meaning of a song and the interpretation of the tone of voice of the singer more strongly than sentence type. Even the compellingness of directives compared to questions or statements, seems to be in no way comparable to the effect of meter.

Please note that this is just about sentence 'type', the effect of the text of each sentence specifically is integrated in the random effect of sentence. Nevertheless, given the relatively strong effect of melody and accompaniment on all ratings, it is clear that music can affect the interpretation of the singer's intentions, his tone of voice, and even his sincerity.

Conclusion

The results of the current experiment only show small effects of timing. Moreover, these effects cannot unambiguously be related to timing as such. The effect of dissonance as a result of shifting the melody in relation to the accompaniment might also play a part. On the other hand, several aspects of the design might have attenuated the effect of timing. Therefore, future research might search for other ways to create somehow ecologically valid, varied, and attractive combinations of melody and accompaniment, that would not cause dissonances (presumably a simple vamp, or a djembe beat); to create a more consistent set of statements to be rated; and/or to create stimuli in which not all

An important secondary finding of this study is that music affects the interpretation of the singer's tone of voice, his state of mind, his intentions and his sincerity, and subsequently will affect the emotional meaning of a song. Music even seems to overrule the effect of an important linguistic factor such as sentence type. As the various aspects of music and language affecting the ratings were just meant to be able to generalize the hypothesized effect of timing, they were not properly counterbalanced and their effects have to be interpreted with caution. Nevertheless, they show the need of research about the way music affects the interpretation of song, and hopefully inspire new experiments. Creating authenticity is very important in popular music, and tone of voice is crucial in music therapy, advertisement and games.

Acknowledgements. I would like to thank my supervisors Emile Wennekes, Frank Hakemulder and Roel Willems, for their comments; my colleagues from UIL OTS for their help with Lime Survey and statistics and giving me the opportunity to present and discuss part of these results in an ELiTu Talk; my colleagues from musicology for their critical comments on the stimuli; Christan Grotenbreg for recording the stimuli and digitally manipulating the recordings; several bookshop owners and web hosts for helping me finding participants; my participants for completing the survey; and NWO, the Dutch Government, and SG het Rhedens for granting me the opportunity to avail myself of a PhD scholarship for teachers.

References

Aaftink, C. (2014) *Kaleidoscope: A phenomenological-empirical study of beauty. University of Alberta* (Doctoral dissertation). Retrieved from: era.library.ualberta/files/1c18dg58c/Aaftink_Cathelein_201409_PhD.pdf.

Ali, S. O. & Z. F. Peynircioğly (2006) Songs and emotions: are lyrics and melodies equal partners. *Psychology of music, 34(4)*, 511-534.

Bouwer, F., Schotanus, Y. P., Sadakata, M., Mullensiefen, D & Schaefer, R. (in preparation) *Measuring musical sophistication in the low countries; validation of a Gold MSI translation in Dutch*.

Brackett, D. (1995) *Interpreting popular music*. Cambridge/New York/Melbourne: Cambridge University Press.

Burns, L. (2000) Analytic methodologies for rock music: harmony and voice leading strategies in Tori Amos's 'Crucify'. Everett, W. (Ed) *Expression in pop-rock music: a collection of critical and analytical essays*. New York/London: Garland.

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, 68-91.

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, 352.

Hakemulder, J.F. (2004). Foregrounding and its effect on readers' perception. *Discourse Processes*, *38*(2), 193-218.

Honing, H. J., Ladinig, O. Winkler I. & Háden G. (2009). Is beat induction innate or learned? Probing emergent meter perception in adults and newborns using event-related brain potentials (ERP). *Annals of the New York Academy of Sciences*, nr. 1169: *The Neurosciences and Music: Disorders and Plasticity*, p. 93-96.

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, 170-177.

Jones, M.R. (1976). Time, our lost dimension: Toward a new theory of perception, attention and memory. *Psychological Review*, 83, 323-3.

Jung, H., Sontag, S., Park, Y.S. & Loui, P. (2015) Rhythmic Effects of Syntax Processing in Music and Language. *Frontiers in Psychology*, *6*, 1762.

Large, E.W. & Jones, M.R. (1999). The dynamics of attending: how people track time varying events. *Psychological Review*, 106, 119–159.

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

Menon, V. & Levitin, D. (2005). The rewards of music listening: Response and physiological connectivity of the mesolymbic system. *NeuroImage*, 28(1), 175-184.

Miall, D. & Kuiken, D. (1994). Foregrounding, Defamiliarization, and Affect Response to Literary Stories. *Poetics*, 22, 389-407.

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). Questionnaire retrieved from: https://www.gold.ac.uk/music-mind-brain/gold-msi/download/.

Pattison, P. (2015) Lesson 46: Phrasing. *Songwriting: writing the lyrics*. Berkelee: Coursera. Retreived, March, 2015. https://www.coursera.org/learn/songwriting-lyrics/lecture/3BtUo/phrasing.

Quené, H. & Port (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica* 62 (1), 1-13.

Scherer, K., Trznadel, S., Fantini, B. & Sundberg, J. (2017). Recognizing emotions in the singing voice. *Pschychomusicology: Music, Mind and Brain*, 27(4), 244-255.

Shanahan, D., & Huron, D. (2014). Heroes and villains: The relationship between pitch tessitura and sociability of operatic characters. *Empirical Musicology Review*, 9, 46-59.

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

Temperley, D. (2001) *The cognition of basic musical structures*. Camebridge, MA/London, UK: MIT.

Appendix

Examples of the remaining eight melodies (Figure 3, for the first one see Figure 1), and the non-closure ratings for each of them, plus the distribution of sentence-type, tempo, meter and measure across melodies (Table 5.)

Non-closure ratings were largely in line with musical theory. A fragment ending on the dominant (220) received high non-closure 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#, also received a very low non-closure rating, even by musically trained listeners, although it ends with an out-of-key chord.

Table 5. Mean non-closure rating, meter, measure, tempo and sentencetype per melody. Features are not counterbalanced.

Melody	Non-clos.	Meter	Tempo	Measure	Sent.[1]
100	0.13	Trochee	Low	4/4	d/s/s
130	0.22	Trochee	Moderate	3/4	d/d/d
150	0.48	Trochee	Low	3/4	d/s/q
170	0.10	Trochee	Moderate	4/4	d/d/q
190	-0.28	Trochee	Low	3/4	d/d/q
210	-0.69	Trochee	Low	3/4	e/e/s
220	0.44	Iamb	High	3/4	d/q/s
250	-0.28	Iamb	High	4/4	d/d/d
280	-0.13	Iamb	Moderate	3/4	q/s/e

[1] Sentence type: d = directive; s= statement; q = question, e = ellipse

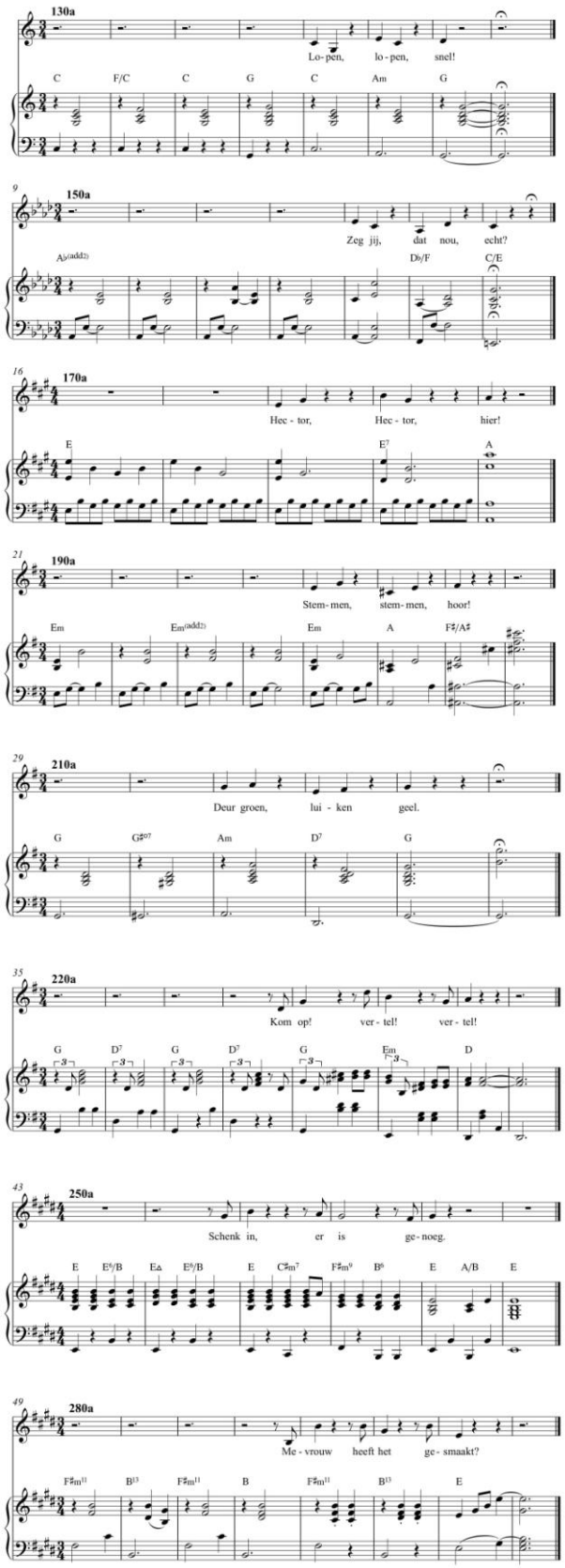

Figure 3. Examples of melodies: 130, 150, 170, 190, 210, 220,250 and 280 in on-beat position. For melody 100, see Figure 1.