

Commentary on Schotanus, “Singing and Accompaniment Support the Processing of Song Lyrics and Change the Lyrics’ Meaning”

CHRISTOPHER S. LEE[1]
Goldsmiths College, University of London, UK

ABSTRACT: In this commentary, a number of problematic aspects of the studies presented in the target article are discussed. Suggestions have been made for further analyses of some of the data and for additional experimental investigations.

Submitted 2020 August 7; accepted 2020 August 9.

Published 2020 October 22; <https://doi.org/10.18061/emr.v15i1-2.7774>

KEYWORDS: *Singing and language, learning, attention, emotion*

THE two studies presented in the target article are part of a program of research (Schotanus, 2015; Schotanus, Eekhof & Willems, 2018) investigating the possible beneficial effects of singing as a pedagogical mode of presentation, which builds on previous work that has found beneficial effects on verbatim memory in language-learning contexts (Ludke, Ferreira & Overy, 2013; Tegge, 2015, and references therein). The studies focus on three claims derived from the author’s general theory (dubbed the Musical Foreground Hypothesis): (i) singing can facilitate the processing of language, with beneficial effects in particular on intelligibility and recall; (ii) these effects can be enhanced by song accompaniments; and (iii) music can add meaning to a song by increasing the emotional impact and meaningfulness of the lyrics. The claims were tested by presenting participants in the two studies with different versions of four unfamiliar cabaret songs (composed by the author) in which the lyrics were sung or spoken (or vocalized to “la”), with or without accompaniment, and asking them to rate the different versions on measures of language processing, emotional content, and other variables of interest. The main predictions were that sung compared to spoken and accompanied compared to unaccompanied versions would yield higher ratings on the relevant measures, and predictions relating to language intelligibility and comprehensibility in particular were (partially) confirmed.

Given the pedagogical goal outlined by the author in the target article, an important strength of the research presented here is the high level of ecological validity of the main study (Study 1), which was conducted with 271 adolescent school pupils in a classroom setting. However, this strength comes at a price, as the range and complexity of the hypotheses being tested, along with some methodological problems, make the experimental findings difficult to interpret and to generalize. In this commentary, I focus first on the methodological issues and then on the problem of generalizing the findings, and suggest further analyses and experimental investigations that might help provide solutions.

METHODOLOGICAL ISSUES

The main methodological issues concern the questionnaire used in the studies and the outcome variables derived via factor analyses from item responses (for all experimental materials and analyses, see Schotanus, 2017). As the questionnaire is a novel instrument that has only been used in the two studies, there are inevitably a number of unanswered questions regarding its use, and there are three issues of particular importance.

The first issue concerns the use of self-evaluative rating questions for text comprehension. It is well known that self-assessments of understanding and of knowledge more generally typically yield varyingly over-inflated estimates (e.g. Lin & Zabrocky, 1998, on reading comprehension; Rozenblit & Keil, 2002, on the more general phenomenon). Asking participants to rate the lyrical texts on the extent to which they are “intelligible” (“goedtebegrijpen”) and “comprehensible” (“goedteverstaan”) is therefore likely to yield less



reliable estimates of their cognitive performance, the outcome of interest, than behavioral tasks such as word identification and text recall (as indeed the author acknowledges), and it is a pity that the only such task that produced usable results was a text recall task in Study 2 (reported in Schotanus et al, 2018).

The second issue concerns the comprehensibility of the rating questions. Several questions in the two versions of the classroom study questionnaire (Study 1) were not understood by some participants (the meaning of “poëtisch”/“poetic” in one of the lyric appreciation questions, “klankherhaling”/“sound repetitions” in a question about rhyme perception, possibly “drammerig”/“nagging” in one of the emotion rating questions, and in the first version the use of “muziek”/“music” to refer to the melody in the questions in the a cappella condition), amounting to 27 out of all 153 questions (17.7%) in the first version and 17 out of all 160 questions (10.6%) in the second version. In the lab study questionnaire (Study 2), these problems were rectified, though a further review process might have uncovered other problematic items (e.g. did participants distinguish “goedtebegrijpen”/“intelligible”, presumably referring to word identification, from “goedteverstaan”/“comprehensible”, presumably referring to meaning?). In short, while the second study may have been largely free of problems, the first study was not, and while only a minority of questions are likely to have caused problems, the resultant increased burden of processing imposed on participants could have led some of them to adopt a “satisficing” (Krosnick, 1991) style of responding, thereby resulting in poorer-quality data all round (Lenzner, 2012).

The third issue concerns the factor analyses. Given that not all questions were asked in all song conditions, a single factor analysis was not possible, and so a different analytic strategy was required. Unfortunately, a different strategy was chosen for each study:

1. In Study 1, separate factor analyses were performed over five groups of questions probing distinct domains that were asked in all the same song conditions (“Processing fluency”, “Voice”, “Lyrics”, “Emotion”, and “Repetition”).
2. In Study 2, factor analyses were performed over the same two groups of questions probing emotional content and repetition (“Emotion” and “Repetition”) as in Study 1, while a single separate factor analysis (reported in more detail in Schotanus et al, 2018) was performed over (most of) the questions asked in the three song conditions where there was text and music (accompanied speech, a cappella, and complete).

The main consequence of this decision is to make it difficult to compare the results of the two studies, because the outcome variables for each study are different. This is true even for the emotional content analyses, as the analysis in Study 1 includes a song condition (accompaniment only) that was not used in Study 2. In particular, it is difficult to know what to make of the divergent results for the Clearness factor (a combination mainly of intelligibility and comprehensibility) in Study 1 and the related Positive Affect factor (a combination of intelligibility, comprehensibility and positive emotions) in Study 2. In Study 1, Clearness yields higher scores in both sung conditions (complete and a cappella) compared to the spoken song conditions, whereas Positive Affect in Study 2 only yields high scores in the complete song condition, with no difference between the a cappella and spoken song conditions. The divergent results could be simply a consequence of the different factor structures, but they could also signal a real difference in experimental outcomes.

It would be possible to go some way towards addressing these issues with a reanalysis of the data. The comparability issue and to an extent the data-quality issue could be addressed by running identical factor analyses over responses to the same sets of questions from the two studies, excluding (where possible) questions that proved problematic in Study 1 and song conditions that were not used in Study 2, and comparing the results in detail side by side (factors extracted, item loadings, and results of the statistical analyses). Addressing the first issue would be more difficult, but in the case of Study 2 it might be possible to gain at least a rough idea of the reliability of the self-evaluative ratings by examining the extent to which they predict performance on the recall task.

GENERALIZABILITY OF THE FINDINGS

A fundamental goal of any study involving experimental research is to produce findings that can be generalized beyond the experimental setting, and in the case of a study aiming to influence pedagogical practice, that goal would include being able to generalize the findings to a variety of different pedagogical situations. However in the case of the present study it is unclear to what extent this is possible, because the

individual effects of the factors putatively influencing outcomes and their likely varying importance across key situational variables (speaker/singer characteristics, music style etc.) largely remain to be determined, and so their varying joint effects in different situations are not easy to predict.

The clearest example of this problem concerns the language processing benefit. Both in the target article and elsewhere (Schotanus, 2015), the author hypothesizes that the benefit arises principally from the effects of music on attention and mood: (i) the metrical structure of the vocal line and accompaniment focuses the listener's attention on important linguistic events (via a dynamic attending process of the sort proposed by Jones and colleagues; e.g. Large & Jones, 1999); (ii) prosodic features of the vocal line (lengthening, increased intensity) enhance the prominence of such events and also help signal phrase structure (Palmer & Kelly, 1992; Palmer & Hutchins, 2006); and (iii) music-induced changes in arousal and mood lead to improved all-round cognitive performance (see Schellenberg & Weiss, 2013, and references therein). However the hypothesis is largely untested: (i) although studies such as Johnson, Huron & Collister (2014) and Gordon, Magne & Large (2011) show that the non-coincidence of metrical and linguistic stress in vocal lines impairs intelligibility, they don't show that their coincidence actually improves intelligibility compared to ordinary speech (see note 7, Gordon et al, 2013); (ii) studies of sung prosody similarly have not compared singing with ordinary speech and so an intelligibility benefit for sung prosodic features compared to those of ordinary speech remains to be demonstrated; and (iii) music-induced changes in arousal and mood have varying effects on different tasks (Schellenberg & Weiss, 2013), and it is not clear how large the effect may be on tasks involving language processing. It would therefore be useful if the two studies here could supply relevant evidence. Unfortunately however they cannot, because the experiments have not been designed to disaggregate the individual contributions of the factors of interest, which likely operate in concert across the different song conditions, with metrical clarity probably increasing from spoken to sung (along perhaps with prosodic prominence) and unaccompanied to accompanied conditions, thereby tracking positive mood (Figure 3, top panel).

The upshot is that the studies deliver a "black box" result that is difficult to generalize: the factors responsible remain undetermined and so the answers to critical questions remain obscure. What would be the likely effect of songs with different levels of metrical complexity and hence metrical clarity? Or of songs which induced different levels of arousal and different moods? What would be the effect of different singing styles? More broadly still would the rhythmic structure of the language make any difference? Would languages like French or Spanish that have a more even syllabic rhythm than Dutch and that tolerate more stress mismatching in their text-settings (Rodriguez-Vazquez, 2010; Temperley & Temperley, 2013) yield a lesser metrical and prosodic benefit from singing?

It would of course be unreasonable to expect a single study to answer all these questions, but if the relative contributions of the different factors could be determined, it would at least give some indications as to likely answers. That could be done with a small-scale carefully controlled lab study, while phonetic analyses of the sung and spoken lyrics used in the two studies here might provide useful additional insights.

CONCLUSIONS

The studies reported in the target article are an ambitious attempt to demonstrate the benefits of singing as a pedagogical mode of presentation. The main strength of the research is the high level of ecological validity of the classroom study, but there are some methodological problems and the experimental hypotheses are not fully explored. However, these issues could be addressed with further analyses and investigations.

ACKNOWLEDGEMENTS

This article was copyedited by Tanushree Agrawal and layout edited by Kelly Jakubowski.

NOTES

[1] Correspondence can be addressed to christopher.lee@gold.ac.uk.

REFERENCES

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>

Johnson, R. B., Huron, D., & Collister, L. (2014). Music and lyrics interactions and their influence on recognition of sung words: an investigation of word frequency, rhyme, metric stress, vocal timbre, melisma, and repetition priming. *Empirical Musicology Review*, 9(1), 2-20. <https://doi.org/10.18061/emr.v9i1.3729>

Krosnick, Jon A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–36. <https://doi.org/10.1002/acp.2350050305>

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

Lenzner, T. (2012). Effects of survey question comprehensibility on response quality. *Field Methods*, 24(4), 409-428. <https://doi.org/10.1177/1525822X12448166>

Lin, L.-M., & Zabrocky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, 23(4), 345–391. <https://doi.org/10.1006/ceps.1998.0972>

Ludke, K. M., Ferreira, F., & Overy, K. (2013). Singing can facilitate foreign language learning. *Memory and Cognition*, 41(5). <https://doi.org/10.3758/s13421-013-0342-5>

Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language*, 31(4), 525–542. [https://doi.org/10.1016/0749-596X\(92\)90027-U](https://doi.org/10.1016/0749-596X(92)90027-U)

Palmer, C., & Hutchins, S. (2006). What is musical prosody? In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 46. The psychology of learning and motivation: Advances in research and theory* (p. 245–278). Elsevier Academic Press. [https://doi.org/10.1016/S0079-7421\(06\)46007-2](https://doi.org/10.1016/S0079-7421(06)46007-2)

Rodriguez-Vazquez, B. (2010). Text setting constraints: A comparative perspective. *Australian Journal of Linguistics*, 30, 19-34. <https://doi.org/10.1080/07268600903134004>

Rozenblit, L., & Keil, F. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, 26, 521-562. https://doi.org/10.1207/s15516709cog2605_1

Schellenberg, E. G., & Weiss, M. W. (2013). Music and cognitive abilities. In D. Deutsch (Ed.), *The psychology of music* (p. 499–550). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-381460-9.00012-2>

Schotanus, Y. P. (2015). The musical foregrounding hypothesis: How music influences the perception of sung language. In Ginsborg, J., Lamont, A., Philips, M. & Bramley, S. (Eds.) *Proceedings of the Ninth*

Triennial Conference of the European Society for the Cognitive Sciences of Music, 17-22 August 2015, Manchester, UK.

Schotanus, Y.P. (2017). Supplemental materials for publications concerning three experiments with four songs, hdl:10411/BZOEEA, Dataverse NL Dataverse, V1.

Schotanus, Y. P., Eckhof, L. S., & Willems, R. M. (2018). Behavioral and neurophysiological effects of singing and accompaniment on the perception and cognition of song. In Parncutt, R. & Sattmann, S. (Eds.). *Proceedings of ICMPC15/ESCOM10*. Graz, Austria: Centre for Systematic Musicology, University of Graz. 389-394.

Tegge, F. A. G. (2015). *Investigating song-based language teaching and its effect on lexical learning*. Unpublished doctoral dissertation, Victoria University of Wellington, New Zealand.

Temperley, N., & Temperley, D. (2013). Stress-meter alignment in French vocal music. *The Journal of the Acoustical Society of America*, 134(1), 520–527. <https://doi.org/10.1121/1.4807566>