The Influence of Interactions Between Music and Lyrics: What Factors Underlie the Intelligibility of Sung Text?

JANE GINSBORG

Centre for Music Performance Research, Royal Northern College of Music, Manchester, UK

ABSTRACT: This document provides a brief commentary on Johnson, Huron and Collister's (2014) article entitled "Music and lyrics interactions and their influence on recognition of sung words: An investigation of word frequency, rhyme, metric stress, vocal timbre and repetition priming." The commentary is written from the point of view of someone who is not only a fellow researcher in the field, but also a former professional singer and singing teacher. It begins by summarizing a previous study by two of the authors and the research of two further teams of investigators in the field. It concludes by focusing on each of the authors' eight hypotheses, findings and discussion in turn.

THE study reported by Johnson, *et al.* (2014) is a follow-up to Collister and Huron's (2008) investigation of the intelligibility of sung text in English. To summarize the background to that study, Smith and Scott (1980) and Benolken and Swanson (1990) reported that listeners find it hard to distinguish between different vowels when sung at high pitch by a trained soprano. Hollien, Mendes-Schwartz and Nielsen (2000) showed that even voice teachers, phoneticians and speech pathology students found it hard to identify vowels correctly when trained singers – male as well as female – sang too high, i.e. when the fundamental frequency of the voice reaches or exceeds the first formant. Collister and Huron reported that, in their study, listeners (neither expert musicians nor singers) had greater difficulty understanding sung text. Listeners were found mishearing more than seven times as many words when compared to spoken text. Three-quarters of the errors involved consonants. Collister and Huron noted that when listeners did not identify vowels correctly, they tended to hear monophthongs as diphthongs or, more often, confuse them with central vowels (for those who are not expert in linguistics, these – in addition to the example Johnson et al. give of hearing "beat" as "bet" – can be found at http://web.ku.edu/~cmed/ipafolder/vowels.html).

A HOT TOPIC

Sung text intelligibility is currently something of a "hot" topic. To declare an interest: I am a member of another research team working in this field. Our starting point was, unsurprisingly, our own experiences of finding it hard to understand sung text, particularly when performed by unamplified opera singers as opposed to those using amplification, as in musical theatre, or performing operettas, such as those by the English Victorians Gilbert and Sullivan without amplification. First, we asked listeners, regardless of their musical background, for their views on the factors underlying intelligibility, or the lack of it (Fine & Ginsborg, 2007a). Next, partly on the basis of our own experience as singers and singing teachers, as well as Hollien et al.'s (2000) findings, we asked other singers and singing teachers for their views (Fine & Ginsborg, 2007b).

Responses to our questionnaire informed the design of two experiments in which we used two groups of listeners, expert and non-experts, and manipulated the number of singers they heard performing meaningful and "scrambled" texts in English (Fine, Ginsborg & Barlow, 2009; Ginsborg, Fine & Barlow, 2011). We found that sung text was more intelligible to singers than non-singers, particularly when sung by a soloist rather than a choir; age and experience of listening was also associated with task performance such that younger participants found texts more intelligible than did older participants. All participants found it easier to make out the words on second hearing, and when the text was meaningful.

Another research team is ploughing a similar furrow. Edward Wickham and The Clerks, an *a cappella* vocal group specialising in Renaissance music, are currently undertaking a project funded by the Wellcome Trust, *Tales of Babel*, combining the performance of specially-composed music and lyrics with research using members of the audience as participants (e.g. Heinrich, Wickham, Fox, Cross & Hawkins, 2013). Their work is informed by the literature on auditory streaming. They are interested in the potential effects of number of voices (in fact there was no effect) and the sex of the singers to whom the audience is attending, or by whom they are distracted (male intelligibility increased when distracters' voices were those of females).

Thus three teams are pursuing the same topic, albeit focusing on different aspects of the problem of intelligibility. Johnson, Huron and Collister are in the lead, however, so far as publication is concerned; they cannot be blamed for being unfamiliar with the other researchers' findings to date. And to a considerable extent their eight hypotheses make sense both intuitively and on the basis of existing evidence, including that presented by Collister and Huron (2008). I will comment on the findings and discussion in relation to each one in turn. The task undertaken by listeners, who were neither expert singers nor phoneticians, was to identify target words from the context of recorded short phrases sung and spoken by two groups of singers: operatic and experienced in musical theatre.

HYPOTHESES, FINDINGS AND DISCUSSION

First, common words were, as predicted, more intelligible than rare words, although the authors point out that the latter are more frequent in sung than spoken texts. They argue that "it is difficult to imagine a situation where archaic words would be more intelligible than common words; unless we were dealing with a very knowledgeable set of listeners who are *very* familiar with musical lyrics containing archaic words." Listeners' experience was one of the variables Fine et al. (2011) manipulated, although we would need to re-analyse our data to establish if archaic words were actually more intelligible than common ones (our stimulus materials included words such as "smithy" and "anvil"; in a scrambled version of the text it would not have been possible to predict the latter from the former). Presumably the listeners in Johnson et al.'s study wrote down the target words, as did the participants in our study. Archaic words, being less familiar, may be harder to spell than common words, so when we analysed our data we agreed on acceptable variants. Perhaps common words are not more intelligible but simply easier to write down.

Second, it was predicted that diphthongs would be less intelligible than monophthongs, but the reverse was found. This is perhaps not surprising since Collister and Huron (2008) noted that a common error made by their participants was hearing monophthongs as diphthongs. Johnson et al. suggest that "sung vowels differ from spoken vowels, so sung pure vowels are more difficult to recognize than their spoken counterparts." Classically-trained singers are taught to use only the five primary or "basic" Italian vowels (see http://vocalcoach.hubpages.com/hub/How-to-Sing-the-Five-Basic-Singing-Vowels for an example of standard advice to singers) in almost all circumstances and in almost all languages, including English, singly or in combination. My guess is that the listener has longer to predict the meaning of the word, in addition to obtaining information from the glide from one vowel to the next. Tangentially, although Johnson and his colleagues did not investigate the role of consonants in the present study, it is worth noting that some singing teachers and, in particular, choral conductors, require singers to adapt these as well, in the belief that they improve intelligibility ("Ph" for "V" at the beginning of a word, or "Ch" for "J", producing phrases such as that well-known delicatessen "Cheeses of Nazareth"). This belief may, however, be mistaken.

The third and fourth hypotheses were both upheld: words set to melismas are less intelligible than syllabic settings, and settings preserving the words' stress when spoken facilitated intelligibility – useful advice for composers! The fifth hypothesis, that repetition increases intelligibility, was supported too, but only for immediate, as opposed to delayed repetition; the authors point out that other text-setting features, such as the melody, may interfere with intelligibility.

The sixth hypothesis combined the fourth and fifth hypotheses, predicting that it is helpful to listeners if a word that has just been sung as a single syllable is followed by the same word set to a melisma. In fact the reverse was found to be the case; it was as though the melisma primed the listener for the single syllable. The authors' post-hoc speculation is that listeners retain the first syllable heard in the melisma even though they do not necessarily recognise the word; this would be congruent with my proposal that, as for Hypothesis 2, the listener simply has longer to predict or infer it from the context.

Similarly, the seventh hypothesis was also disconfirmed: rhyme significantly decreased, rather than increased intelligibility. The authors attribute this finding to listeners' use of memory when identifying target words. Copeland and Radvansky (2001) describe a phonological similarity decrement, first identified by Baddeley and Dale (1966) and confirmed more recently by Lobley, Baddeley and Gathercole (2005), such that phonologically distinct words are better recalled than phonologically similar words, in the context of lists. Rhyme is a form of phonological similarity, so this is one plausible explanation. Another explanation, however, is also presented: participants may have found it hard to distinguish rhyming words from repeated words, which occurred frequently – as they often do in real song lyrics.

The final hypothesis derives from the findings of Smith and Scott (1980) and the common observation that the relative importance of lyrics is different in different genres; it was predicted that the words sung by singers experienced in musical theatre would be more intelligible than words sung by opera singers. This too proved not to be the case. The authors put forward the explanation that too few singers were used to record the stimulus materials and suggest that further research on different vocal styles is warranted (they cite Lithuanian dialects as examples of contrasting diction).

I agree on both counts. In our not-dissimilar research we used singers with a wide variety of vocal characteristics. While, in the study described above, our intention was only to vary the number of singers used to present the stimulus materials, we could have compared them on a variety of dimensions including experience (e.g. of solo / choral singing; with / without amplification; in different genres), use of vibrato, and vocal timbre. None of these alone predicts the degree to which singers are intelligible, since their motivation – and indeed determination – to be understood must also be important factors. In short, vocal "personality" is predicted by more than the singer's training. Future researchers would therefore be wise to have decided in advance and to specify the characteristics desired for each comparison, if distinctions are to be made between styles of singing. However, even this may not be enough: as we have seen, and as the present authors acknowledge, listeners' experiences and expectations of singers in different contexts – and, indeed, different environments, both physical and acoustic – also have a role to play.

CONCLUSION

While Johnson *et al.* (2014) observe in their conclusion that only half their predictions were upheld, and three findings were statistically significantly in the wrong direction, as it were, there is in my view no need to be apologetic about this. As they say, it was an exploratory study, and for this reason they justify establishing an alpha level of 0.1 for their statistical tests. All the hypotheses were based on previous findings and/or the experience of music-lovers accustomed to listening to vocal music. The findings in relation to the mishearing of vowels (and for that matter consonants) came as less of a surprise to me, when I reflect on my own experience as a singer, and I would argue – on the basis of other listeners' views – that an even larger range of factors potentially underlie the intelligibility of sung text than the authors suggest. I look forward very much to their future reports of further studies undertaken in this field

ACKNOWLEDGEMENT

REFERENCES

Baddeley, A. D., & Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior, 5*(5), 417-420.

Benolken, M. S., & Swanson, C. E. (1990). The effect of pitch-related changes on the perception of sung vowels. *Journal of the Acoustical Society of America, 87*(4), 1781-1785.

Collister, L., & Huron, D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review*, *3*(3), 109-125.

Copeland, D., & Radvansky, G. (2001). Phonological similarity in working memory. *Memory & Cognition*, *29*(5), 774-776.

Fine, P., & Ginsborg, J. (2007a). Perceived factors affecting the intelligibility of sung text. In K. Maimets-Volk, R. Parncutt, M. Marin & J. Ross (Eds.) *Proceedings of the third Conference on Interdisciplinary Musicology* (CIM07). Tallinn, Estonia, 15-19 August 2007, http://www-gewi.uni-graz.at/cim07/.

Fine, P., & Ginsborg, J. (2007b). How singers influence the understanding of sung text. In A. Williamon & D. Coimbra (Eds.), *Proceedings of the International Symposium on Performance Science*, European Association of Conservatoires (AEC), Utrecht, The Netherlands. ISBN 978-90-90224-84-8.

Fine, P., Ginsborg, J., & Barlow, C. (2009). The influence of listeners' singing experience and the number of singers on the understanding of sung text. In A. Williamon, S. Pretty & Ralph Buck, *Proceedings of the International Symposium on Performance Science 2009*, European Association of Conservatoires (AEC), Utrecht, The Netherlands. ISBN 978-94-90306-01-4.

Ginsborg, J., Fine, P., & Barlow, C. (2011). Have we made ourselves clear? Singers' and non-singers' perceptions of the intelligibility of sung text. In: A. Williamon, D. Edwards & L. Bartel, *Proceedings of the International Symposium on Performance Science 2011*, European Association of Conservatoires (AEC), Utrecht, The Netherlands. ISBN 978-94-90306-02-1.

Heinrich, A., Wickham, E., Fox, C., Cross, I., & Hawkins, S. (2013). Stream segregation of speech in live concert-hall performances given by a 6-voice choir. Available at http://www.talesfrombabel.co.uk/wp-content/uploads/2013/02/Heinrich_etal_BSA.pdf.

Hollien, H., Mendes-Schwartz, A. P., & Nielsen, K. (2000). Perceptual confusions of high-pitched sung vowels. *Journal of Voice*, *14*(2), 287-298.

Johnson, R., Huron, D., Collister, L. (2014). Music and lyrics interaction 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.

Lobley, K., Baddeley, A., & Gathercole, S. (2005). Phonological similarity effects in verbal complex span. *Quarterly Journal of Experimental Psychology*, *58A*(8), 1462-1478.

Smith, L., & Scott, B. (1980). Increasing the intelligibility of sung vowels. *Journal of the Acoustical Society of America*, *67*(5), 1795-1797.