

# Have we made ourselves clear? Singers and non-singers' perceptions of the intelligibility of sung text

**Jane Ginsborg<sup>1</sup>, Philip Fine<sup>2</sup>, and Christopher Barlow<sup>3</sup>**

<sup>1</sup> Centre for Music Performance Research, Royal Northern College of Music, UK

<sup>2</sup> Department of Psychology, University of Buckingham, UK

<sup>3</sup> Faculty of Technology, Southampton Solent University, UK

The intelligibility of sung text is an important component of listeners' enjoyment of vocal music and a central concern for singers and, for example, choral conductors. Expert listeners such as singers and singing teachers may be better than non-singers at perceiving sung text. We replicated and extended an earlier study investigating the intelligibility of semantically and non-semantically meaningful words performed solo and by a small group of trained soloists by carrying out an experiment in which we manipulated listening expertise, type of text, number of singers, and time of hearing. Participants listened twice to four songs with meaningful and "scrambled" lyrics, sung in unison by a choir and solo, and wrote down as many of the words as they could discern. Singers were better at the task than non-singers; more words were recorded on the second hearing and when the words were meaningful. Sung text involves distortions of consonants and vowels to which singers may be more accustomed, so that they find it easier to discern texts even when scrambled. Choirs may be harder to understand than soloists because their phonemes are more variable and less clear. In future research we will use operatically trained soloists and polyphonic choirs.

*Keywords:* choral; expertise; lyrics; solo; vocal

The intelligibility of sung text is an important component of listeners' enjoyment of vocal music and a central concern for singers and, for example, choral conductors (Fine and Ginsborg 2007a). Factors underlying intelligibility include performer (e.g. number, vocal technique) and listener attributes. It may be easier for one singer to convey the words of a song, and their mean-

ing, than for a group of singers. Vocal technique relates to the culture of Western classical music. “Trained” singers learn to project their voices and use clear diction (solo singers: Adams 1998, Falkner 1994; choral singers: Emmons and Chase 2006). The modification of vowels in the interests of preserving the musical line (Hollien *et al.* 2000) and consonant confusions (Collister and Huron 2008) can affect intelligibility, however, as can the use of vibrato (Sundberg 1994)—all of which might be thought of as distorting the “natural” attributes of the lyrics. The singer, however, can only do so much to ensure intelligibility (Fine and Ginsborg 2007b); the perception of sung text depends to a certain extent on the listener. Those who are themselves experienced singers and singing teachers, and are therefore members of the same culture as the performer, are more likely to be attuned to factors affecting singers’ diction, and better than non-singers at resolving the acoustic signal into recognizable words. In a preliminary experiment, listeners with experience of both singing and listening to singing wrote down significantly more of the words of songs—whether sung by a trained soloist or a small group of trained soloists—than listeners without such experience (Fine *et al.* 2009).

We have replicated and extended Fine *et al.*’s (2009) study, first, by replacing the small group of trained soloists with an unaccompanied chamber choir, to test more realistically the difference between the intelligibility of a soloist and a group of choral singers. Second, while we compared once again listeners’ ability to discern the words of semantically- and non-semantically-meaningful songs, we constructed the latter using an improved strategy.

## METHOD

### Participants

Twenty-four singers (7 male, 17 female), aged 19 to 81 years (median=24.5), were recruited from a music college and an amateur choir; 24 non-singers (18 male, 6 female), aged 18 to 62 years (median=20.0), were recruited from students and staff in a university department. The singers reported a mean of 26.8 (SD=23.6) years’ experience of singing, 6.9 (SD=6.7) hours’ singing, and 5.1 (SD=4.6) hours’ active listening to singing in the seven days prior to taking part in the experiment, while equivalents for the non-singers were 2.1 (SD=3.5) hours’ singing and 5.1 (SD=10.8) active listening to music generally.

### Materials

A comparison was made of the ability of singers and non-singers to discern and write down, on first and second hearings, the words of songs (“meaning-

ful” vs. “scrambled”) when sung by a soloist and a choir. Four short, novel, songs were performed *a cappella* by a solo soprano and a choir singing in unison with the correct lyrics in English and with the same lyrics scrambled. All stimuli were recorded in the same room at 48 kHz, 24 bit resolution onto a digital audio workstation, and then encoded as MP3 files at a constant bit rate of 320 kbit/s using a Neumann KM130 omnidirectional condenser microphone placed about 1.3 m from the singer(s). Stimuli were played to participants on a laptop computer using either its internal speakers or external speakers. All participants stated that the stimuli were loud enough.

### **Procedure**

In the first part of the study each participant was required to provide demographic information and then complete a daily singing and listening diary for seven days either in response to daily e-mails or on paper. The second part of the study consisted of the experiment. Each participant was tested individually. The researcher gave the participant a piece of paper and two pens of different colors. Having explained the procedure, the researcher played a sequence of short songs to the participant. Each song was played twice. The participant was encouraged to write down the words using one of the two pens while listening to the recording. The first time the song was played the recording was stopped at the end of each of the four lines of text. The second time it was played through without stopping, and the participant was asked to use the other pen to indicate alterations or additions to the words s/he heard.

*Analyses:* We used the informational semantic match (ISM) method described by Hustad (2006) for transcribing speakers with dysarthria. The number of syllables (including phonemes) transcribed acceptably was calculated as a percentage of the maximum possible. Credit was given for misheard consonants that were feasible in the context of the preceding or next word (e.g. “and choose” for “and shoes”) but not mis-heard vowels unless very close. Word order, morphological, segmentation, and spelling errors were ignored, as were “additional” words not in the target material. Two of the researchers scored a proportion of the data independently, and agreed in 93% of syllables transcribed. Disagreements were resolved following discussion.

*Scoring:* It was not possible to test all participants in all conditions, so a repeated-measures ANOVA could not be undertaken. Instead, a linear mixed model analysis was carried out using SPSS. The dependent variable was percentage of syllables transcribed acceptably. There were four fixed factors: group (singer vs. non-singer), number of singers (solo vs. choir), words (meaningful vs. scrambled), and attempt (first vs. second). Correlations were

Table 1. Main effects.

		Mean %	$F_{44,132}$
Participants (listeners)	Singers	67.0 (SD=20.9)	
	Non-singers	55.8 (SD=23.8)	9.2**
Performance	First	56.9 (SD=23.2)	
	Second	65.9 (SD=22.1)	36.2***
Text	Meaningful	76.3 (SD=17.7)	
	Scrambled	46.5 (SD=17.4)	398.0***
Performer(s)	Soloist	65.6 (SD=22.0)	
	Choir	57.1 (SD=23.3)	33.1***

Table 2. Bivariate correlations between demographic variables and task performance.

Participants	Singers	Non-singers	All
Age	R=-0.548**	R=-0.437*	R=-0.349*
Years' experience	R=-0.544**	NS	NS
Hours' singing	NS	NS	R=0.328*
Hours' listening	NS	NS	Rho=0.388**

Note. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

obtained between the percentage of syllables transcribed acceptably and the participant's age, experience of singing (in years, generally, and hours during the previous week, specifically), and recent active listening to singing. Exploratory data analysis revealed one set of outlying scores in each group. When these scores were removed temporarily, the significance of the results was not altered, so results, means, and standard deviations (SDs) are reported for all participants.

## RESULTS

As shown in Table 1, there were significant main effects of all four variables such that singers found sung text more intelligible than did non-singers. The soloist was more intelligible than the choir. Generally, it was easier to make out the words on second hearing and when the text was meaningful.

As shown in Table 2, younger participants performed better on the task than older; there were also significant positive correlations between hours of singing and hours of listening per week and task performance.

## DISCUSSION

As hypothesized, singers found sung text more intelligible than did non-singers, supporting the conclusion that the sharing of a culture in which it is acceptable to distort consonants and vowels, as they are normally produced in the context of speech, can compensate for the “unnaturalness” of such modifications and the use of vibrato. Generally, it was easier to make out the words on second hearing than first and when the words were meaningful than when they were scrambled; audiences would benefit if composers and singers alike took note of this finding.

The results also support the hypothesis that the same lyrics sung by a soloist are likely to be more intelligible than those sung by a choir. If the comprehension of sung text, like speech, relies on the clarity of consonants and vowels, choirs may be harder to understand because they produce more variable and less clear phonemes. It is also possible that intelligibility, in this study, was affected by pitch (Hollien *et al.* 2000, Di Carlo 2007). While the soloist was not operatically-trained, she was a soprano singing in a high register (for her); the choral performances may have been more intelligible because the tenors and basses—although singing in unison—were singing an octave and two octaves below the sopranos. This will be addressed in a future study by comparing the intelligibility of high and low soloists. Also, the extent to which soloists are more intelligible than choirs is likely to depend on ability and training. It may be, for example, that operatically-trained soloists are more intelligible than singers without such training.

Finally, the finding that task performance apparently deteriorated with age and experience may be a function of the relative non-homogeneity of the sample of listeners; while two-thirds of the singers were students, the remainder were the considerably older members of an amateur choral society, while only a small group of non-singers was recruited from members of staff in their 50s and 60s. This too will be addressed in future studies.

### Acknowledgments

We are very grateful to the singers who created the stimulus materials, all the participants, our research assistants Polly Long (Southampton Solent) and Jasper Brownrigg (Royal Northern College of Music), and Eddie Shoesmith (University of Buckingham) for his advice on the use of linear mixed model analysis.

### Address for correspondence

Jane Ginsborg, Centre for Music Performance Research, Royal Northern College of Music, 124 Oxford Road, Manchester M13 9RD, UK; *Email*: jane.ginsborg@rncm.ac.uk

### References

- Adams D. (1998). *A Handbook of Diction for Singers: Italian, German, French*. Oxford: Oxford University Press.
- Collister L. B. and Huron D. (2008). Comparison of word intelligibility in spoken and sung phrases. *Empirical Musicology Review*, 3, pp. 109-125.
- Di Carlo N. S. (2007). Effect of multifactorial constraints on intelligibility of opera singing (II). *Journal of Singing*, 63, pp. 550-567.
- Emmons S. and Chase C. (2006). *Prescriptions for Choral Excellence: Tone, Text, Dynamic Leadership*. Oxford: Oxford University Press.
- Falkner K. (1994) *Voice*. London: Kahn and Averill.
- Fine P. and Ginsborg J. (2007a). Perceived factors affecting the intelligibility of sung text. In K. Maimets-Volk, R. Parncutt *et al.* (eds.). *Online Proceedings of the Third Conference on Interdisciplinary Musicology (CIM07)*, accessed at [www-gewi.uni-graz.at/cim07](http://www-gewi.uni-graz.at/cim07).
- Fine P. and Ginsborg J. (2007b). How singers influence the understanding of sung text. In A. Williamon and D. Coimbra (eds.), *Proceedings of ISPS 2007* (pp. 253-258). Utrecht, The Netherlands: European Association of Conservatoires (AEC).
- Fine, P., Ginsborg, J., and 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, and R. Buck (eds.), *Proceedings of ISPS 2009* (pp. 51-56). Utrecht, The Netherlands: European Association of Conservatoires (AEC).
- Hollien H., Mendes-Schwartz A. P., and Nielsen K. (2000). Perceptual confusions of high-pitched sung vowels. *Journal of Voice*, 14, pp. 287-298.
- Hustad K. C. (2006). A closer look at transcription intelligibility for speakers with dysarthria: Evaluation of scoring paradigms and linguistic errors made by listeners. *American Journal of Speech-Language Pathology*, 15, pp. 268-277.
- Sundberg J. (1994). Acoustic and psychacoustic aspects of vocal vibrato. *KTH Quarterly Progress and Status Report*, 35, pp. 45-68.