Intelligibility of spoken and sung sentences in different types of background noise

Sarah Knight[a], Sarah Hawkins[b], Edward Wickham[b], Christopher Fox[c] and Antje Heinrich[a]
[a]MRC Institute of Hearing Research, Nottingham, UK; [b]University of Cambridge, UK;
[c]Brunel University, London, UK

Supported by wellcome trust

Background

- Investigations of speech perception in noise suggest:
  - Lower **signal-to-noise ratios** (SNRs) are more detrimental than higher SNRs [1].
  - **Multi-talker babble** is a more effective masker than noise [2].
  - **Semantic context** aids intelligibility [3].
- However, such investigations **have not been extended to the perception of singing**, despite the frequency with which we encounter sung words in everyday life.
- Singing, as opposed to speech, presents unique and additional challenges to intelligibility:
  - Musical rhythm is generally prioritised above speech rhythm, thus **disrupting temporal information** at the syllable level which may aid segmentation and lexical access [4].
  - Vowels are pitched, thus **distorting spectro-temporal information** carried in formants which may aid detection of phonetic content [5].
  - Listening to sung text may be regarded as an **aesthetic and/or musical** – as opposed to information-bearing – activity, thus potentially negating a benefit of semantic context.

Question: Can the findings of SPiN research regarding SNR, background type, and semantic context be extended to the perception of sung text?

Stimuli and Method

- 36 **high- and low-predictability context** sentence pairs modelled after the SPIN-R sentences [6], e.g.:
  High predictability: *"They borrowed money to pay the school fees".*
  Low predictability: *"Deb only offered to get the full fees".*
- 3 **noise types**: shifting vowel sounds, spoken babble, /sh/ + silence.
- 2 **noise levels**: high and low.
- Data collected as part of 6 **LIVE** concerts given by a professional British choir (The Clerks).
- Sentences sung by a male singer.
- Noise performed live by other choir members.
- **Design fully crossed** across the 6 concerts.
- Audience members chose the **final word** from four options (correct response + three foils) using handheld devices.
- **4AFC** (4-alternative forced choice) task.
- The foils were:
  1. highly **phonetically similar** to the target but semantically implausible
  2. highly **semantically plausible** (with respect to the high-predictability context) but phonetically different
  3. moderately **phonetically similar** and **semantically plausible**
- 354 participants.

Hypotheses

NOISE TYPE	Silence	/sh/	shifting vowels	babble
COMMENTS	Control condition.	Similar to high-pass steady-state noise: energetic masking, but mostly in non-relevant frequency regions.	Similar to speech-modulated sound: energetic masking in relevant frequency regions; pitch shifting may cause informational masking	Rich background: energetic and informational masking.
HYPOTHESISED PERFORMANCE	Close to ceiling.	< silence	< /sh/	< vowels

- With respect to **noise level**, performance is predicted to be worse in the high level condition (low SNR) than the low level condition (high SNR), since the former causes greater energetic masking [1].
- With respect to **predictability**, performance is predicted to be worse for low-predictability sentences than high-predictability sentences [3], if listeners process sung speech similar to spoken speech. If sung speech is not listened to for information-bearing purposes, semantic predictability of context may be inconsequential to intelligibility.

Results and Discussion

Figure 1 shows the main effects of noise type, predictability and noise level, all of which were in the directions predicted and were highly significant overall ($p < 0.001$*).

Figure 2 shows the significant interactions, both of which were highly significant overall ($p < 0.005$*). The interaction between noise type and predictability (Fig. 2A) indicates that a significant effect of semantic context was apparent for all conditions except silence. Moreover, the interaction remained significant ($p < 0.005$*) even when the silence condition was removed from the model, indicating that the effect of context was not comparable across noise conditions.

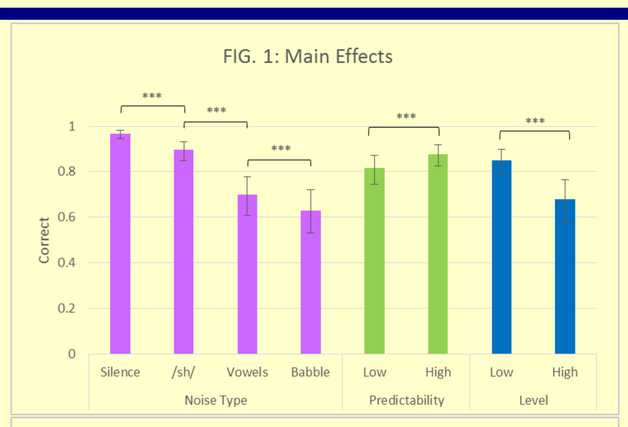
The interaction between noise type and noise level (Fig. 2B) indicates that the effect of level was non-significant in the /sh/ condition, but highly significant for all other noise types. Possible reasons for the non-significant level effect for /sh/ include a ceiling effect or an inability of the singers to produce the /sh/ sound at a sufficiently high level.

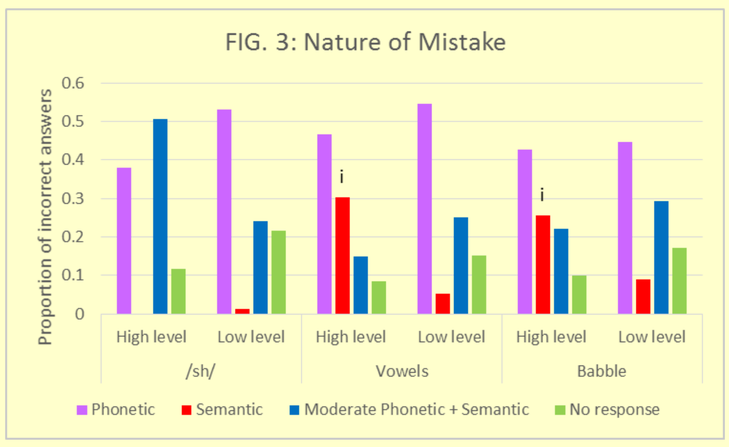
Since the foils were chosen according to strict criteria, it is possible to examine the nature of participants' mistakes. Figure 3 shows the nature of the mistakes for the intelligibility conditions depicted in Fig. 2B.

*Generalized linear mixed model. Fixed effects = noise type, level, predictability. Random effects = concert, participant.

** $p < 0.05$; pairwise comparison, LSD adjusted.

*** $p < 0.001$; pairwise comparison, LSD adjusted.





In most conditions participants tended to choose the highly phonetically similar foil when making a mistake, suggesting that some phonetic information was available to them. The semantically plausible but phonetically dissimilar foil (labelled "i" in Fig. 3) tended to be chosen most often in the listening conditions most challenging to intelligibility: vowel or babble backgrounds in high levels of background noise ("ii" in Fig. 2B). This could be interpreted as indicating an increased reliance on semantic context when access to phonetic detail is reduced. However, if this were the case, one might expect to see the largest effect of semantic context on accuracy in these challenging conditions – whereas the effect of context on accuracy was in fact relatively small, at least in the vowel condition ("iii" in Fig. 2A). Further work is needed to clarify this observation.

Conclusions

- This study sought to extend our understanding of variables affecting speech intelligibility to singing.
- It sought to use an **ecologically valid approach** by collecting data during live concerts, in a range of venues, and testing a broad range of audience members. To enable data collection, multiple choice responses and condition testing across concerts were used.
- Despite all these deviations from laboratory-led research typical of SPiN research and despite acoustic differences between sung and spoken speech, initial analyses suggest that **many of the findings from SPiN research replicated to singing**, including:
  - a detrimental effect of background noise specific to the type of background.
  - An effect of **background noise level.**
  - a benefit of **semantic context** (although possibly smaller than for spoken speech)**.**
- The materials are currently being validated in a laboratory setting.

References

[1] Tun, P. A. & Wingfield, A. (1999). One voice too many: Adult age differences in language processing with different types of distracting sounds. *The Journals of Gerontology, 54B (5)*, pp. P317-P327.
[2] Simpson, S & Cooke, M. P. (2005). Consonant identification in *N*-talker babble is a nonmonotonic function of *N*. *The Journal of the Acoustical Society of America 118*, pp. 2775-2778.
[3] Pichora-Fuller, M. K. (2008). Use of supportive context by younger and older adult listeners: Balancing bottom-up and top-down information processing. *International Journal of Audiology 47 (S2)*, pp. S72-S82.
[4] Cutler, A. & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language 31 (2)*, pp. 218-236.
[5] Carlsson, G. & Sundberg, J. (1992). Formant frequency tuning in singing. *Journal of Voice 6 (30)*, pp. 256-260.
[6] Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M. & Rzeczkowski, C. (1984). Standardization of a test of speech perception in noise. *Journal of Speech, Language and Hearing Research*, *27 (1)*, pp. 32-48.

Acknowledgements

Many thanks to The Clerks, Mark Edmondson-Jones, the concert venues and promoters, and all our audience members.