

Commentary on “Comparison of Word Intelligibility in Spoken and Sung Phrases” by Lauren Collister and David Huron.

JOHAN SUNDBERG[1]

*Department of Speech, Music and Hearing
School of Computer Science and Communication*

STEN TERNSTRÖM

*Department of Speech, Music and Hearing
School of Computer Science and Communication*

ABSTRACT: We note that the intelligibility of vowels is reduced at high pitches, and propose that decreased decoding of vowels may partially account for the reported findings. Analysis of performance as a function of pitch is recommended. We surmise that reverberation associated with the large microphone distance might have generated reverberation that interfered disproportionately with identification of consonants in sung stimuli, which usually have greater sound intensity than spoken stimuli. Finally, we note the potential relevance of vibrato for speech intelligibility.

Submitted 2008 November 18; accepted 2008 November 25.

THE article is impressive both with regard to quality of presentation and quantity of the material analyzed. It is a classical observation that vowel intelligibility in singing deteriorates dramatically at high pitches, and two females and one tenor produced material for the listening test in the investigation. As a consequence, the test words appeared in a rather high pitch range.

A compilation of earlier studies of vowel intelligibility as a function of fundamental frequency suggested that difficulties start at about C5 and develop to a disaster at F5, see Figure. Probably the pitch could account for some of the listener's depressingly frequent failures to hear the correct phonemes in the sung versions, 334 errors versus 46 in the spoken stimuli. The frequent repetitions of the text segments so commonly occurring in classical vocal music seems a smart trick. In any event, the influence of pitch on the present results might have been interesting to see.

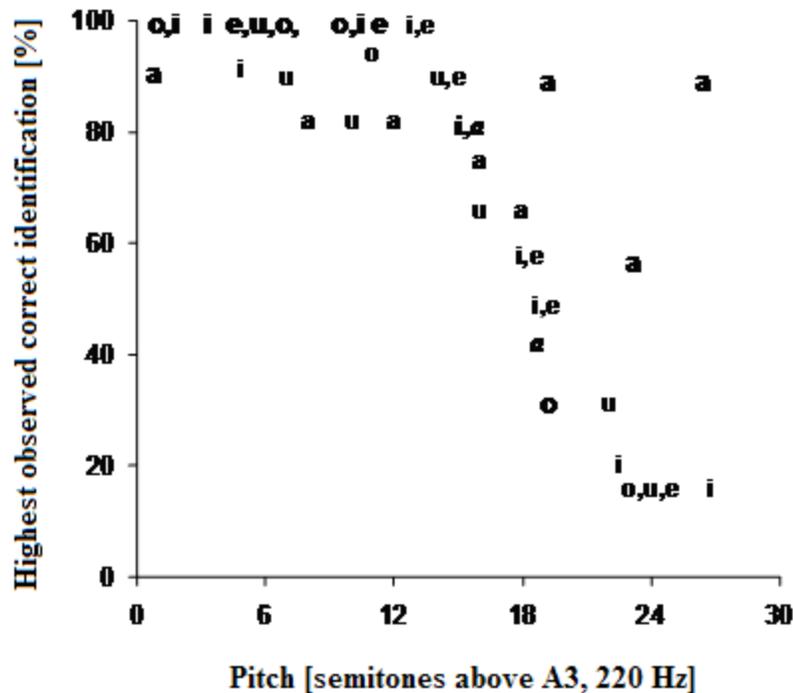


Fig. 1. Figure showing the highest vowel identification percentages observed in various investigations plotted as a function of the fundamental frequency.

Smith and Scott (1980) showed, although not in a strictly controlled experiment, that consonants substantially helped listeners to a correct vowel identification at pitches above F5. This in turn suggests that the formant transitions are crucial to vowel intelligibility at high pitches. It is possible that Collister and Huron ran into some bias as all their test words appeared in an identical phonetic context, following an alveolar version of the consonant /d/.

The microphone distance of 10 m is rather large. Although no details of the auditorium acoustics are given, the diffuse field is likely to have dominated the sound in the recorded stimuli, physically if not perceptually. Firstly, then, it is likely that the intelligibility of both spoken and sung tokens would have been higher with stimuli recorded in a smaller room. Secondly, the sung version will generally be louder than the spoken version and in particular on vowels, generating a stronger diffuse field on vowels. Hence it is probable that in the sung case, consonants are more likely to be masked by the reverberation and thus consonant intelligibility will be lower. This of course is part of the ecology of the situation, but the effect may be worth noting. The stimuli were appropriately presented stereophonically over headphones, but it is not stated whether the stereo microphones were arranged explicitly for binaural reproduction, which would have been preferable in this context.

There is no mention in the paper of the amount of vibrato used by the singers, but the paradigm appears to have been classical, so we might assume that normal classical vibrato was used. Vibrato can be expected to improve the intelligibility of high tones, or worsen it, depending on how rapidly the vowel spectrum envelope changes in relation to the reverberation of the room.

On a related issue, van Besouw et al (2005) found, among other things, that listeners retained less of the linguistic information when phrases were sung rather than spoken.

Even though the listening test included a very great number of stimuli there are some older references, scattered in various publications that may complement the understanding of text intelligibility in singing:

NOTES

[1] Department of Speech, Music and Hearing, Kungliga Tekniska högskolan (Royal Institute of Technology), SE-100 44 Stockholm, Sweden.

REFERENCES

van Besouw, R., Howard, D. M., & Ternström, S. (2005). Towards an understanding of speech and song perception. *Logopedics Phoniatics Vocology*, Vol. 30, No. 3-4, 129-135.

Morozov V P (1965) Intelligibility in singing as a function of fundamental voice pitch, *Soviet Physics Acoustics* Vol. 10, 279- 83.

Nelson H D & Tiffany W R (1968) The intellegibility of song, *National Association of Teachers of Singing Bulletin Dec.*, 22-33.

Scotto di Carlo N & Germain A. A perceptual study of the influence of pitch on the intelligibility of sung vowels, *Phonetica*, Vol. 42, 188.197.

Sundberg J & Gauffin J (1982) Amplitude of the voice source fundamental and the intelligibility of super pitch vowels, i R Carlson & B Granström, utg., *The Representation of Speech in the Peripheral Auditory System*, New York: Elsevier Biomedical Press, 223-228.

Sundberg J, Romedahl C Text Intelligibility and the Singer's Formant—A Relationship? In press, *Journal of Voice*.

Sundberg J, Vibrato and vowel indentification, *Archives of Acoustics* 2 (1977) 257-266.