Spectral distribution of solo voice and accompaniment in pop music

Daniel Zangger Borch[1] and Johan Sundberg

1. Stockholm Music Conservatory, Box 17 528 , SE-118 91 Stockholm

Abstract

Singers performing in popular styles of music mostly rely on feedback provided by monitor loudspeakers on the stage. The highest sound level that these loudspeakers can provide without feedback noise is often too low to be heard over the ambient sound level on the stage. Long-term-average spectra of some orchestral accompaniments typically used in pop music are compared with those of classical symphonic orchestras. In loud pop accompaniment, the sound level difference between 0.5 and 2.5 kHz is similar to that of a Wagner orchestra. Long-term-average spectra of pop singers' voices showed no signs of a singer's formant but a peak near 3.5 kHz. It is suggested that pop singers' difficulties to hear their own voices may be reduced if the frequency range 3-4 kHz is boosted in the monitor sound.

Introduction

In solo singing it is obviously important that the singer's voice can be readily heard against the background of the orchestra. In classical singing, the singer's formant has been found to be an elegant solution to this problem. The singer produces strong partials in a frequency range where even a loud orchestral accompaniment generates partials of comparatively low amplitude (Sundberg, 1987). In singing of popular music, the situation is different. A loud accompaniment is typically provided by synthesizers, electrically amplified instruments and a drum section. As the singers use microphones, the audibility of the singer's voice is generally taken care of by the sound engineer.

To achieve an appropriate degree of phonatory control, singers need to hear their own voices. In pop music, this is often difficult or even impossible because of the elevated sound level of the accompanying instruments. Normally, special monitor loudspeakers placed on the stage, are used to provide this feedback. A problem frequently encountered in live concert situations is that the possibilities to raise the volume of the singer's voice in the monitor system are limited by the risk of generating feedback noise or scream. In-ear systems recently have been introduced as an alternative, which, however, are associated with certain problems. If applied binaurally in-ear phones reduce the singer's auditory contact with the fellow musicians, and this limits the possibilities to live improvisation and exchange of information between and during performances. A monaurally applied in-ear phone creates a highly unnatural sound image. Pop singers often tend to use extreme degrees of vocal loudness and phonatory press when they perform. One reason may be that otherwise they cannot hear their own voice. In other words, the masking effect of the accompanying instruments may deprive them of their auditory proprioceptive feedback. This masking is determined by the sound level and the spectrum of the accompaniment. Thus, pop singers' risk of voice damage seems related to these sound characteristics.

Given this background, it seemed worthwhile to investigate the spectral distribution of solo voices as well as of various kinds of accompaniment typically used in pop music. The investigation was limited to a male singer. The findings will be discussed with regard to the possibilities to improve the audibility of male solo singers' voices without jeopardizing their vocal health and personal voice timbre.

Method

The investigation focuses on the vocal style commonly used in the so-called main stream popular music area, henceforth pop music. Four different sets of recordings were made: (1) a live recording of a pop concert with a male soloist performing at the "Fasching" music house in Stockholm; (2) studio recordings of a musical excerpt as performed by the same pop singer and an opera singer; (3) studio recordings of

Speech, Music and Hearing, KTH, Stockholm, Sweden
TMH-QPSR Volume 43: 31-35, 2002

Table 1. Instruments used in the different accompaniment sounds analyzed.

No.	Instruments playing
1	Drums, Electrical bass, Percussion, Electrical guitar with distorted sound played single string, Tenor saxophone and Digital piano.
2	Drums, Electrical bass, Electrical guitar with distorted sound played single-string, Digital piano.
3	Drums, Electrical bass, Electrical guitar with distorted sound played with power chords, Tenor saxophone, Percussion
4	Drums, Electrical bass, Electrical guitar, Digital piano, Percussion, Tenor saxophone.
5	Drums, Electrical bass, Electrical guitar with distorted sound played with power chords, Tenor saxophone.
6	Drums, Electrical bass, Percussion, Electrical guitar, Tenor saxophone.

four different pop singers' performances of the solo voice of different songs from the pop repertoire; and (4) copies of excerpts from commercial recordings of classical orchestral music by WA Mozart, L Beethoven, P Tschaikovsky, and R Wagner.

Six excerpts representing different examples of common pop music styles were selected from the "Fasching" material and digitized (Table 1).

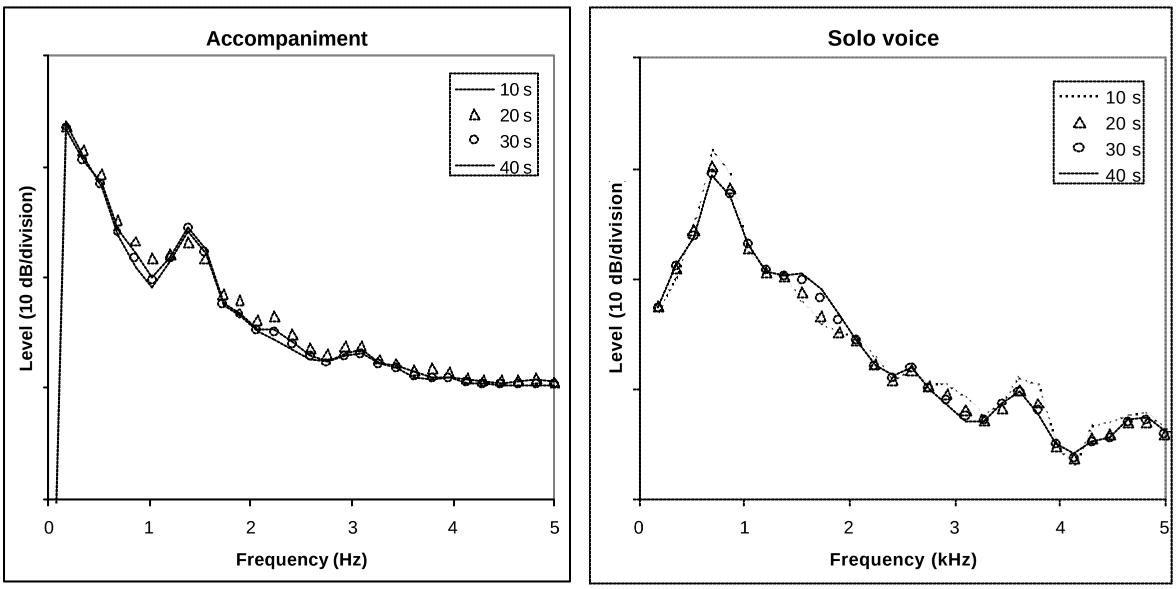
All excerpts were examined by means of long-term-average spectrum (LTAS) analysis (Jansson & Sundberg, 1976a and b), using the Soundswell workstation program package (Ternström, 1992). As the upper limit of the pitch range typically reaches A4, the analysis bandwidth was 400 Hz.

The reproducibility of LTAS was tested by comparing analyses of 20, 30 and 40 s duration of singing and a typical example of accompaniment. The results show that for singing and also for orchestral sound, a stable LTAS was obtained after about 20 s (Figure 1). The LTAS of instrumental music, however, varied considerably depending on the combination of instruments played, as expected.

Results

Figure 2 compares LTAS of the pop singer and an operatic tenor who sang the same excerpt in the same key, duration 17 s, approximately. Both singers used tempo and phrasing of their own choice. In the graph, both spectra were normalized such that the highest LTAS level was set to zero dB. The greatest difference between these two voices concerned the singer's formant.

In the case of the operatic singer, the level below the singer's formant peak is lower than in the case of the pop singer. This would reflect a higher frequency of the third formant combined with a lower frequency of the second formant in front vowels. A high third formant adds to the



Figur 1. Reproducibility of LTAS as demonstrated by analyses of 20, 30 and 40 s duration of singing and a typical example of pop music accompaniment.

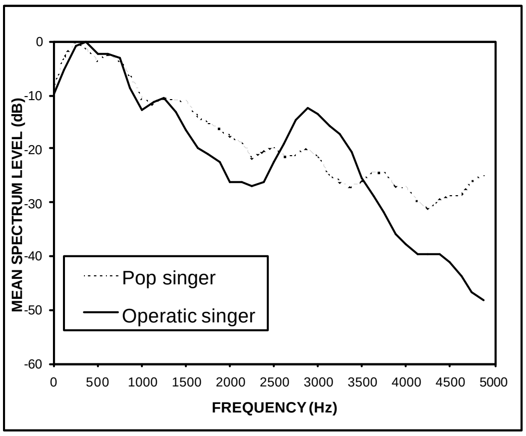


Figure 2. LTAS of a pop singer and an operatic tenor singer who sang the same excerpt in the same key.

prominence of the singer's formant peak and a low second formant in front vowels reduces the vowel dependence of the singer's formant peak and thus contributes to timbral equalization of vowels (Sundberg, 2001). The pop singer showed a considerably higher level near 3.5 kHz than the opera singer. This would be due to higher frequencies of the fourth and fifth formants.

The same pop singer's LTAS can be compared with the LTAS of four other pop singers in Figure 3. The curves vary within a range of about ±3 dB between 2.5 and 4 kHz. None of the curve exhibits a peak similar to the tenor's singer's formant. Rather, the curves are characterized by a set of two or three, less prominent peaks and a more continuous overall decrease of level with increasing frequency. Also, all curves show a peak somewhere between 3.5 and 4.5 kHz. The figure shows that

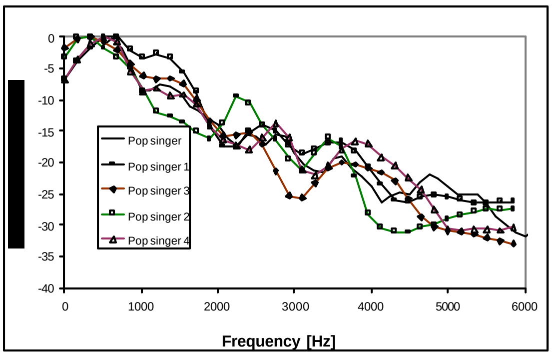


Figure 3. LTAS of the pop singer shown in Fig. 2 and of four other pop singers performing songs of more than 60 s duration.

all these pop singers lacked a singer's formant. Furthermore, LTAS of the pop singer shown in Figure 2 does not deviate from those of the other voices in any remarkable way. This indicates that this singer was reasonably representative for pop singers in general.

The singer's formant has been found to increase the perceptibility of a singer's voice in the presence of a loud orchestral accompaniment; a long-term average spectrum of a classical symphonic orchestra typically shows a peak near 0.5 kHz and a rather steep slope above this frequency (Sundberg, 2001). This implies that the singer's formant is located in a frequency region where the competition from the accompaniment is moderate. In a sense, then, the operatic singing voice and the accompaniment seem to be mutually optimized. Hence, it is interesting to examine long-term average spectra of accompaniments that are typically used in pop music.

Figure 4a shows LTAS of a variety of instrumentations typically occurring in pop music, see Table 1. For the sake of comparison, the levels have been normalized with respect to the highest level, which mostly occurred in the vicinity of 100 Hz. In the case of example 4, a peak appeared near 0.6 kHz, which was produced by a solo tenor saxophone. The mean relative level at 3 kHz varied considerably between the examples. Instrumentations 5 and 2 presented the highest and lowest relative levels in this frequency region, amounting to –25 dB and –44 dB, respectively.

For comparison, Figure 4b shows LTAS of five excerpts from commercial recordings of loud sections from classical music played by symphonic orchestras. The levels are normalized as in Figure 4a. In these cases, high levels occurred up to about 400 Hz, and above this frequency the LTAS contour sloped rather steeply toward higher frequencies. The relative level near 3 kHz is similar to those found for the pop music examples.

The normalization of the curves in Figure 4 depended on a peak in the bass range near 100 Hz. The sound level at these low frequencies are, however, not likely to have a great influence on the masking of a singer soloist's voice. Of greater interest would be the mean LTAS slope above 0.5 kHz. Table 2 compares the average LTAS slope in the frequency range 1-5 kHz. For the pop ensembles, the mean slope varies between –7 and –12 dB while for the classical examples the mean slope varies between –13

Speech, Music and Hearing, KTH, Stockholm, Sweden
TMH-QPSR Volume 43: 31-35, 2002

Table 2. Mean LTAS slope in the frequency range of .50 – 5 kHz for the indicated styles and ensembles.

Pop accomp. no.	**Slope dB/oct**	**Icpt dB**	**r**	**Symphonic Orchestras Example**	**Slope dB/oc**	**Icpt dB**	**r**
1	-7.5	60.8	0.982	R Wagner: Vorspiel	-12.8	136.64	0.985
2	-4.4	14.8	0.925	P Tshaikowski. Conc	- 8.4	91.225	0.950
3	-7.1	85.8	0.978	L Beethoven Egmont	-10.7	116.98	0.981
4	-7.1	58.0	0.968	L Beethoven symph	- 9.6	100.43	0.964
5	-5.5	45.9	0.979	WA Mozart symph	-11.8	116.41	0.962
6	-4.7	23.5	0.960				
Mean SD	-6.0	48.1	0.965		-10.4	111.3	0.970
	1.4	26.1	0.021		1.9	20.0	0.016

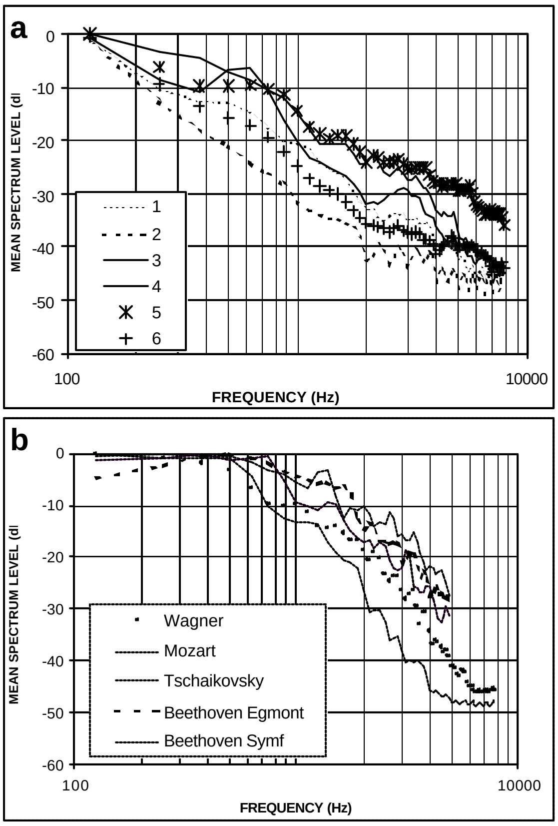


Figure 4a. LTAS examples of different types of pop orchestras. Numbers refer to the ensemble playing as specified in Table 1.

Figure 4b. LTAS of excerpts from loud orchestral music by the indicated composers.

and –20 dB. The slopes, averaged across examples, are –9.7 dB and –16,7 dB for the pop ensembles and the classical orchestras, respectively. This indicates that average spectrum slope between .5 and 5 kHz is considerably less in pop ensembles than in classical symphonic orchestras.

Discussion and Conclusions

As mentioned in the introduction an important aspect of singing with loud accompaniment is the solo singers' possibilities to hear their own voices. Our LTAS measurements have shown some interesting similarities between the sound level characteristics of the classical orchestral accompaniment and the accompaniment frequently used in pop music. In the rock example of pop music accompaniments, the relative sound level near 3 kHz was comparable to that occurring in a Wagner orchestra. It seems reasonable to assume that Wagner singers can hear their own voices even when the accompaniment is loud. If so, the results appear to suggest that solo singer's chances to hear their own voices in the pop music genre would be improved if they sang with a singer's formant. This solution, however, is not in accordance with vocal traditions accepted within these styles of music. A better possibility would be to raise the level in the 3-4 kHz frequency range in the monitor sound that provides the singers with auditory feedback of their voices.

Our comparison between operatic and pop singers' long-term-average spectra was based on recordings of two singers only. The most salient difference was the presence and absence of a singer's formant, respectively. With regard to the opera singer, there is no reason to doubt the generality of the observation; a singer's formant is notoriously observed in male operatic singer voices. With regard to our pop singer his LTAS was similar to those of other pop singers. In a recent investigation of country singers' voices, it was found that also these singers lack a singer's formant peak (Cleveland et al., 2001). This suggests that singing without a singer's formant is practiced also in other types of popular music.

An LTAS peak near 3.5 kHz has been found in voices characterized as "good" (Leino & Kärrkäinen, 1995; Leino, 1994; Nawka et al., 1997). Likewise, Cleveland and collaborators (2001) found such a peak in their country singers both when they were singing and when they were speaking. This appears to suggest that pop singers use a "good" speaking voice when they sing. On the other hand, no information on pop singers' speaking voice is available.

Monitors on the stage are sometimes replaced by earplug phones in the soloist's ears. Even though ear plug phones are associated with some disadvantages, e.g., difficulties to hear direct spoken communication on the stage during a performance, they are quite wide-spread. They eliminate problems with feedback noise and screams and possibly also reduces the threat of damaging both hearing and voice caused by high sound levels on the stage.

This study was a first survey of the acoustic working conditions in pop singing. Another important aspect that should be relevant to vocal pedagogy is what phonatory characteristics successful pop singers use. We intend to return to such aspects in a future investigation.

Acknowledgements

This is a revised version of the first author's thesis work in music pedagogy at the Royal College of Music, Stockholm. The support of a grant from the Patricia Gramming Memorial Foundation and from the Stockholm Music Conservatory is gratefully acknowledged.

References

Cleveland T, Sundberg J & Stone RE (2001). Long-term-average spectrum characteristics of country singers during speaking and singing. *J Voice* 15: 54-60.

Jansson EV & Sundberg J (1976a). Long time average spectra applied to analysis of music. Part 1: Method and general application. *Acustica* 34: 15-19.

Jansson EV & Sundberg J (1976b). Long time average spectra applied to analysis of music. Part 2: An analysis of organ stops. *Acustica* 34: 275-280.

Leino T & Kärrkäinen P (1995). On the effects on vocal training on the speaking voice quality of male student actors. *Proc of Intl Conf of Phonetic Sciences* 1995/3: 496-499.

Leino T (1994). Long-term-average spectrum study of speaking voice quality in male actors. In: Friberg A, Iwarsson J, Jansson E & Sundberg J, eds, *SMAC 93 (Proc of the Stockholm Music Acoustics Conf, 1993)*, Royal Swedish Academy of Music, 79: 206-210.

Nawka T, Anders LC, Cebulla M & Zurakowski D (1997). The speaker's formant in male voices. *J Voice* 11: 422-428.

Sundberg J (2001). Level and center frequency of the singer's formant, *J Voice* 15/2: 176-186.

Sundberg J (1987). *The Science of the Singing Voice.* DeKalb, Illinois: Northern Illinois University Press.

Ternström S (1992). Sound Swell Signal Work-station, software manual.