

Balancing spectra between different speaking styles

Gunilla C. Thunberg

Department of Linguistics, Stockholm University

For some languages enhancement of the energy in the upper part of the speech spectrum has been shown to contribute to perception of emphasis or prosodic focus. Changes in the spectral balance may also be employed for non-linguistic purposes. Classically trained singers make use of a *singer's formant*, that will enable them to vocally compete with an orchestral accompaniment. Only a few attempts have been made to identify any corresponding *speaker's formant*. The work presented here aims at gaining insight into this area of speech research.

1. Introduction

The concept of *spectral balance* is fairly young within speech research. During the past decade the linguistic use of energy enhancement in the upper part of the speech spectrum has been empirically explored (Sluijter & van Heuven, 1993, 1996; Campbell, 1995; Sluijter, Shattuck-Hufnagel, Stevens, & van Heuven, 1995; Sluijter, van Heuven & Pacilly, 1997; Heldner, 2003). It has been shown that spectral balance may be employed for linguistic purposes at the word or phrase level (often in interaction with other features such as *duration* or *intonational* gestures). However, changes in the spectral energy balance may be employed also for non-linguistic or extralinguistic purposes. When utilized over longer stretches of speech, where a whole utterance or even an entire monologue might be characterized by consistent enhancement of the upper parts of the spectrum, this feature would be regarded as being indicative of *voice quality* rather than having a direct linguistic function.

1.1. *The singer's formant and the speaker's formant*

The notion of spectral balance is rarely mentioned in descriptions of voice quality, although the phenomenon has been observed in research on singing already in 1934, by Bartholomew. *The singer's formant*, on the other hand, is a well established concept among both voice and music acoustics researchers (Sundberg, 1974, 1987; Schultz-Coulon, Battmer & Riechers, 1979; Gauffin & Sundberg, 1989; Titze & Story, 1997). Roughly explained, it is produced by clustering the formants 3-5. Thus, a great amplitudinal peak is formed around 3kHz, which, from a perceptual point of view, makes it easier for the singer to vocally "compete" with an orchestral accompaniment. This strategy adheres to opera performances by trained classical singers, although it does not apply for voices with relatively high fundamental frequencies. The effect is achieved mainly by lowering the larynx which will expand the epi-laryngeal tube, thus also entailing an expansion of the vocal tract. Subglottal pressure also plays an important role. Furthermore, there is a direct relationship between the occurrence of a singer's formant and a specific phonatory mode with a steeper slope in the closing phase of the glottal wave (Gauffin & Sundberg, 1989; Alku, Vintturi & Vilkmán, 1999).

Linguistic demands will pose greater constraints on speech production than on singing, making it more complicated to predict the behavior of a corresponding *speaker's formant*. A few attempts have been made to empirically assess some spoken counterpart to the singer's formant, sometimes based on an assumption of a possible transfer from singer-formant strategy into the spoken language produced by the same individual. Professional singers are compared to non-professionals, in singing as well as speaking or reading-aloud tasks. Differences between the two categories are usually found in speaking/reading-aloud as well as in singing. However, an observed transfer of a *singer's*-formant strategy into the spoken language is by no means proof of the existence of a *speaker's* formant. It rather suggests that a professionally trained singer may have adopted a vocal strategy for singing purposes, that could have an impact on his/her production of spoken language as well.

There are also other differences between singing and speaking: singing requires that phonation be maintained throughout; pitch variations, phrasing as well as rhythmical structuring is highly restricted. Manipulation of the spectral balance may be the only strategy available for the singer, in order to obtain a 'carrying voice'. In naturally spoken language you may freely use pitch as well as rhythmical features to convey both linguistic and extralinguistic information to the listener – even voiceless effects may be employed for communicative purposes. Such a freedom is of course relative, as we are subject to quite heavy linguistic constraints; grammatical, and phonological demands must not be compromised, or else our message will not get through. Sluijter & van Heuven (1993, 1996); Campbell (1995); Sluijter et al. (1995); Sluijter et al. (1997); Heldner (2003) have shown that spectral balance may be used for linguistic purposes, although with different strategies between languages. Also, variations in speaking style may reflect the way spectral emphasis would interact with a feature such as overall intensity, signalling prosodic focus, e.g. in Swedish (Heldner, 2003).

The aim of the present study is to explore the use of spectral balance in different speaking styles produced by professionally trained speakers (actors) uttering longer stretches of speech. We would expect to find an enhancement of energy in the upper part of the speech spectrum in a speaking style appropriate for speech in public, produced without any external means for amplification. A professional speaker may employ a different speaking strategy, and even a different technique for production of public speech, as compared to normal conversational speech. Nowadays the use of microphones even in the dramatic theaters is becoming regular, but similar demands on speech have been met without any electronic amplification over thousands of years before us. *Demosthenes* (380-322 B.C.), for example, trying to make his voice heard through the noise of the roaring sea may have been one of the earliest speech researchers actually exploring the production of a *speaker's formant*.

2. Method

The choice of professional actors was made for a number of reasons. Actors must be able to master an array of different voice and speech varieties, mainly because such traits are part of the tools they use for portraying a role character. Furthermore, an actor also has to be sensitive to the *varying productional demands* adherent to acoustic characteristics of different rooms, as well as to the differing conditions for speech produced in different media (on the air, on screen, and on stage). Therefore there is a need for *varying speech techniques* depending on the variety of external conditions. The different techniques employed for speech production under varying conditions will be referred to as *different speaking styles*.

Spoken language produced on a theater stage deals with two seemingly conflicting demands: it should be powerful enough to be perceived by everyone in the audience – still, it must be able to maintain its character of *naturally spoken* language. In order to be distinct

and, at the same time, being able to handle subtle speech nuances, an actor will develop a technique serving as a tool for communicative enhancement. We believe that technique to be the closest we could come to obtaining such a thing as a speaker's formant.

The speakers here are two professional female actors, around the age of forty, both having received professional training at (different) actor's colleges. The text was chosen to be fairly neutral, in order not to generate any overly dramatic interpretation. We chose a short monologue which served well as a text given on stage as well as at the more personal conversational level. The recordings were made on a theater stage. The actors were performing individually, under rehearsal-like conditions. They were wearing a microphone which enabled them to move around freely. The speech signal was registered by a DAT recorder. First, the actor was asked to perform the text by using the speaking style she would ordinarily use on stage. For this part the experimenter (the author) was seated far away from the stage, posing as a receiving target for the spoken text. Next, the experimenter joined the actor on stage, and the actor was asked to perform a shorter (3 min) part of the text, using the conversational speaking style one normally uses at close range.

3. Results

Preliminary results for the two speakers were obtained by means of LTAS, Long-Term Average Spectrum analysis, in order to detect any tendency toward the formation of some kind of speaker's formant in any, or both of the speaking styles. LTAS was applied to portions of speech, by varying lengths. Since there are variations between different parts of the speech material, it would be unwise to completely rely on the averaged measures for longer portions of speech, without taking any local deviations into consideration.

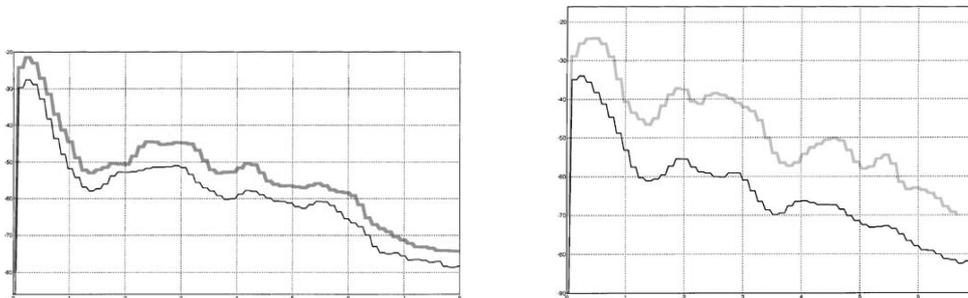


Figure 1. LTAS, Long-Term Average Spectral differences between the two speaking styles as well as between individuals, for the two female speakers; Subject B, left panel, and Subject G, right panel. (NB: since the scaling of the diagrams has not been properly adjusted, you are asked not to pay any attention to absolute levels. These diagrams merely serve as exemplification of different strategies employed in the production of acoustically different speaking styles).

4. Discussion

The results must be regarded separately for each subject. They behave differently, with respect to their fundamental frequency utilization and their formant levels, as well as in the way they differentiate between the two speaking styles. Subject B has higher f_0 values altogether; her fundamental frequency distribution is different from subject G, and the two women appear to be using quite different speaking strategies on the whole. While subject G

might be compared to an alto singer, as far as her vocal register is concerned, subject B would be more of a soprano. According to musical voice research (Sundberg, 1974, 1987) sopranos do not make use of a singer's formant, due to their very high f_0 and a subsequent scarceness in their overtones. They seem to employ some other strategy to obtain a carrying voice. This may explain the different strategies displayed in the results of our two subjects.

Subject G shows a much greater difference in overall intensity between the two styles; as well as via more than one extra peak in the higher frequencies, not only between 2-3 kHz (in the same range as the singer's formant), but in two additional frequency ranges: one at 4-5 kHz, and one at 5-6 kHz. These findings are intriguing, even if there is no risk for conflict with any linguistic demands at those levels.

Further analyses will be made on this material, in order to reveal possible conflicts between the need for a generally enhanced speaking style, on the one hand, and appropriate linguistic signalling, on the other.

5. References

- Alku, P., Vintturi, J. & Vilkmann, E. (1999) Linearity of the function between the sound pressure level of speech and the negative peak amplitude of the differentiated glottal flow for voices of different intensities. *Proceedings of XIVth ICPhS*, San Francisco, 977-980.
- Bartholomew, W.T. (1934) A Physical Definition of "Good Voice-Quality" in the Male Voice, *Journal of the Acoustical Society of America* 6, 25-33.
- Campbell, N. (1995) Loudness, spectral tilt, and perceived prominence in dialogues. *Proceedings of the XIIIth ICPhS*, Dept of Speech Communication and Music Acoustics, KTH, and Dept of Linguistics, Stockholm University, 676-679.
- Gauffin, J. & Sundberg, J. (1989) Spectral correlates of glottal voice source waveform characteristics. *Journal of Speech and Hearing Research*, vol. 32, 556-565.
- Heldner, M. (2003) On the reliability of overall intensity and spectral emphasis as acoustic correlates of focal accents in Swedish. *Journal of Phonetics*, vol. 31, 39-62.
- Schultz-Coulon, H.-J., Battmer, R.-D. & Riechers, H. (1979) Der 3-kHz-Formant – ein Mass für die Tragfähigkeit der Stimme? I. Die untrainierte Normalstimme, II. Die trainierte Singstimme. *Folia Phoniatrica*, vol. 31, 291-301, 302-313.
- Sluijter, A.M.C. & van Heuven, V.J. (1993) Perceptual cues of linguistic stress: intensity revisited. *Proceedings of ESCA Workshop on Prosody*, Lund University, 246-249.
- Sluijter, A.M.C. & van Heuven, V.J. (1996) Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100(4), 2471-2485.
- Sluijter, A.M.C., van Heuven, V.J. & Pacilly, J.J.A. (1997) Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoust. Society of America*. 101(1), 503-513.
- Sluijter, A.M.C., Shattuck-Hufnagel, S., Stevens, K.N. & van Heuven, V.J. (1995) Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English. *Proceedings of the XIIIth ICPhS*, Dept of Speech Communication and Music Acoustics, KTH, and Dept of Linguistics, Stockholm University, 630-633.
- Sundberg, J. (1974) Articulatory interpretation of the "singing formant". *Journal of the Acoustical Society of America*. 55(4), 838-844.
- Sundberg, J. (1987) *The science of the singing voice*. DeKalb, Illinois: N. Illinois Univ.Press.
- Thunberg, G. (1991) Studies of some phonetic characteristics of speech on stage. *Proceedings of the XIIth ICPhS*, Université de Provence, Aix-en-Provence, 5:26-29.
- Titze, I.R. & Story, B.H. (1997) Acoustic interactions of the voice source with the lower vocal tract. *Journal of the Acoustical Society of America*. 101(4), 2234-2243.