TOWARDS THE MEASUREMENT OF THE ACTOR'S FORMANT IN FEMALE VOICES

Sanya Tayal[1], Simon Stone[2] and Peter Birkholz[2]

[1]PEC University of Technology, [2]Technische Universität Dresden
sanyatayal24@gmail.com

Abstract: The Actor's Formant in professional male actors has been a subject of research for many years. It is a strong energy peak that lies in the frequency band between 3-4 kHz in the long-term average spectrum (LTAS) of a professional male actor's voice. Results of perceptual analysis have shown that this concentration of energy has a correlation with the perceived voice quality. While the Actor's Formant has been extensively studied in male subjects, there is little research pertaining to the Actor's Formant in female voices. The aim of this study is to show evidence of the formation of an Actor's Formant in the voices of females as a result of a short vocal warm-up exercise. LTAS results of six female subjects are analyzed and a significant increase in magnitudes of their spectra after the vocal warm up is observed in the region of 4-5 kHz. The study also proposes an experimental method to measure vocal tract transfer functions using transcutaneous excitation of vocal tract. The resulting transfer functions of the subjects are plausible and encourage further studies of formants in female actors and the general population as a whole using the proposed method.

1 Introduction

The inherent difference between the voice of an actor and that of a regular person has been explained by the existence of the Actor's Formant. The Actor's Formant is a strong energy peak that lies in the frequency band between 3-4 kHz in the long-term average spectrum (LTAS) of a male actor's (or, more generally, a professional male speaker's) voice [1]. For a professional singer, a similar formant cluster at a lower frequency in the LTAS (2-3 KHz) is known as the Singer's Formant. Results of perceptual analysis have shown that this concentration of energy has a correlation with the perceived voice quality. Although the Singer's Formant can primarily be achieved through professional training, the Actor's Formant has been found to exist even in untrained, perceptually pleasant male voices [1].

Leino *et al.* stated in [1] that the prominence of an Actor's Formant can be increased by performing specific vocal exercises. He studied the immediate effects of vocal exercises on the voice of a Finnish male actor and found that the LTAS showed a stronger peak in the Actor's Formant region. The exercises aim at producing the best vocal effects in terms of narrowing of the epilaryngeal region. Among other ways of achieving this, phonating into a straw has been proved to be an effective method to establish an Actor's Formant cluster [7, 8].

While the Actor's Formant has been extensively studied in male subjects, its existence for female subjects has not been conclusively demonstrated. Master *et al.* [4] did not find any evidence of an Actor's Formant in her study comparing LTAS of voices of a group of actresses and non-actresses. Here, the participants did not perform any form of specific vocal exercises. However, she reported that Leino did find a peak at 4.3 KHz in both the groups, which was more pronounced in the case of actresses. Unfortunately, Leino never made these results publicly available.

The method of calculating the LTAS is a very convenient and valuable tool for analyzing running speech and singing [3]. It provides information on the spectral distribution of a speech signal over a period of time [2]. In the LTAS, the short term variations in a speech signal are averaged out and the resulting spectrum can be used to study the voice source and the resonance or formant characteristics of the voice [2, 3]. In case of the actor's voice, these characteristics are seen in the form of a high energy peak known as the Actor's Formant which has been known to correlate with good voice quality [1].

To get a more detailed picture of vocal tract characteristics of an actor, the acoustic transfer functions of the vocal tract can be measured for sustained production of vowels. To that end, Badin [5] used transcutaneous excitation of the vocal tract through a white noise signal played by a small loudspeaker placed against the neck (just above the larynx). The vocal tract in this case is assumed to behave as a linear acoustic filter. The broadband nature of the excitation results in an output signal, which has a spectrum that is very similar to the vocal tract transfer function. For such a setup, minimization of the cross-talk between the loudspeaker and the microphone is a major challenge.

In this paper, we present an experimental setup that builds on [5] to measure the acoustic transfer function and improves the setup in terms of crosstalk and reproducibility. The study also aims to show evidence of the existence of the Actor's Formant in female voices as a result of a short vocal warm up. For this, we conducted an experiment with seven female volunteers who included theater actors and professional radio speakers.

2 Materials and Methods

2.1 Measurement System

An experimental setup was developed to measure female vocal tract transfer functions (Figure 1). The setup uses a powerful 3 W and 8 Ω loudspeaker (VECO 34KC08-1). A frequency sweep (150-5000 Hz) of 5 s duration is fed into a 12 W amplifier and played by the loudspeaker. The amplifier was adjusted to give the maximum output power and it was ensured that no clipping occurred.

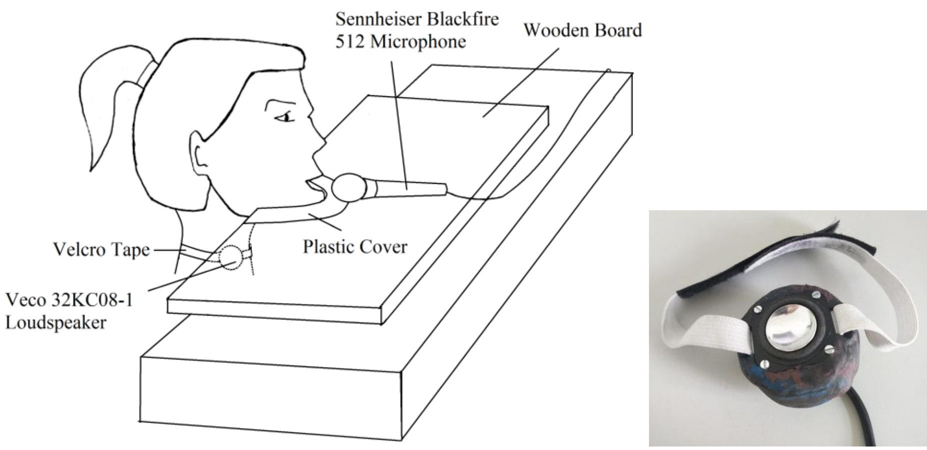

Figure 1 (Left) The proposed experimental setup. The loudspeaker excites the vocal tract near the subject's larynx with an exponential sweep signal. The wooden board is placed firmly on a desk acting as an acoustic barrier to minimize the cross-talk. The subject sustains articulation of a vowel for the duration of the signal. (Right) Magnified view of the loudspeaker in a box acoustically sealed with clay.

The loudspeaker is tied to and pressed against the neck of the subject for transcutaneous excitation and the output pressure signal is recorded through an omni-directional Sennheiser Blackfire 512 microphone kept at a distance of 1 cm from the lips. The major design challenge for the setup was the minimization of the cross-talk between the loudspeaker and the microphone. For this, a wooden board is used as an acoustic barrier between the source at the neck and the mouth opening. The idea of using the wooden board as sound isolation is based on the setup used by Pham Thi Ngoc in [6] that used a board of glass wool as the barrier. The interface between the wood and the chin is padded with a soft plastic for convenience and to cover air spaces through which noise could leak. Also, the loudspeaker is placed in a plastic box filled with sound absorbing paper and is acoustically sealed with clay (Figure 1, right). Figure 2 shows the effect of using the cross-talk reducing wooden barrier and the clay seal of the loudspeaker box on the measurements.

During the experiment, the subject was required to silently articulate a vowel while a frequency sweep signal with duration of 5 s was played through the loudspeaker. The subject voiced the vowel before the measurement to ensure the desired configuration of the vocal tract. To improve the reproducibility of the results, a pulse train is played before the measurements to help determine a suitable position of the excitation source on the neck. This is done by manually adjusting the loudspeaker position until the radiated signal at the subject's lips clearly sounds like the articulated vowel. The subject then voices the vowel again and stops on command after which the sweep signal is played and the articulation is sustained for duration of the signal. This procedure made sure that the configuration of the speaker's vocal tract changed as little as possible.

All the measurements were made with the software 'Measure Transfer Function', which is developed at the Institute of Acoustics and Speech Communication, Technische Universität Dresden. The software provides an inbuilt pulse train and frequency sweep signal of the desired frequency range and amplitude. The software records the signal radiated at the lips, calculates and displays the vocal tract transfer function (Figure 2).

For the measurement of the transfer functions, it was ensured that the cross-talk and external noise sources interfered with the required output signal as little as possible. The setup was tested before the actual measurements were made to check the noise level by performing a measurement with the mouth closed. The breath was held for the duration of the sweep signal. The cross-talk and noise suppression reduced the amplitude of the recorded signal in this closed-mouth setting to below 5 % of the microphone's amplitude range. The impact of the cross-talk and noise suppression on the measurements in the open-mouth setting is shown in Figure 2.

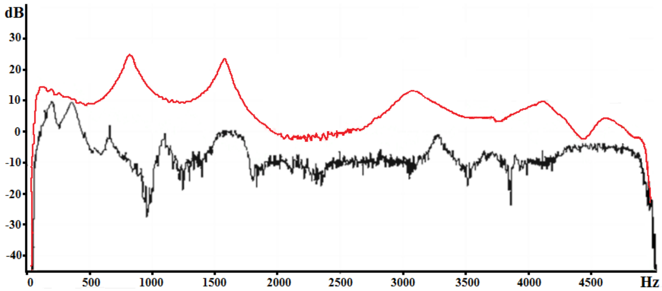

Figure 2 Measured transfer function of vowel [a:] before (black) and after (red) using a wooden board and clay around the loudspeaker to reduce the cross-talk in the experimental setup. Also, better articulation of vowels and the use of a pulse train led to improved transfer functions. The difference in intensities was a result of the variation of the microphone sensitivity. The measurements were made with the in-house software "Measure Transfer Function".

2.2 Subjects and Tasks

The presented study included seven female German native speakers between 22-48 years of age with at least five years of experience as a professional speaker. The experiment was performed in a soundproof studio. The recordings were done with Sennheiser PC-26 headset.

The experiment consisted of two sets of 30 minutes with a break of 15 minutes in between, during which a short vocal warm-up was performed. The subjects were required to use their normal ‘non-stage’ voice for the first set and their ‘stage’ voice for the second set. The subjects performed two tasks in each set. For the first task the subjects had to read aloud a one minute long text (from the German fable ‘Der Nordwind und die Sonne’) before and after the vocal warm-up. The second task was to articulate nine German vowel sounds (/a:, e:, i:, o:, u:, ə, ɛ:, ø:, y:/), of which the transfer functions were measured according to the method described above.

The warm-up exercises were chosen specifically for actors to help in the formation of an Actor’s Formant. They were inspired by Laukkanen *et al.* [7] who elicited a vocal tract adjustment by making subjects phonate vowels into a straw of small diameter. The exercises used in the present study also included performing pitch glides into a straw as shown in the video by Titze [8].

2.3 Analysis

The recordings of the text read by the subjects were used to calculate the LTAS in MATLAB and the differences in the magnitudes of the spectra before and after the vocal warm-up were examined. Silent parts of the recordings were excluded from the analysis by removing the sections below a suitable threshold. The remaining audio data were root mean square (RMS) normalized before the LTAS analysis was performed. To calculate the LTAS, the magnitude of the Fast Fourier Transform (FFT) was calculated with an individual window length of 1024 samples and no overlap between the windows. The FFT of all frames were then averaged to obtain the final LTAS.

3 Results

3.1 LTAS Results

LTAS results of only six out of the seven subjects were analyzed because the last recording was not usable due to technical difficulties. The LTAS results show a considerable increase in energy starting from approximately 4 kHz for all the six subjects. Figure 3 shows the average of the LTAS results of six female subjects before and after the vocal warm up.

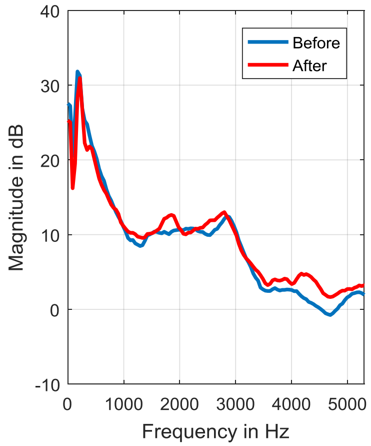

Figure 3 Average of the LTAS of six female subjects before and after the vocal warm up. There is a significant energy increase in the region between 4-5 kHz, indicative of the so-called Actor’s Formant.

The LTAS results for each subject were also analyzed individually. Table 1 and 2 show the mean magnitudes in the ranges of 3-4 kHz and 4-5 kHz. The differences in the mean values before and after the vocal warm-up were generally larger for the range of 4-5 kHz. In addition to that, for the range of 3-4 kHz, 4 subjects showed a significant increase in magnitude (with $p < 0.01$, see Table 1) while 5 out of 6 subjects exhibited a significant increase in the range of 4-5 kHz (Table 2).

Table 1 - Mean magnitudes in the range of 3-4 kHz of LTAS results of six subjects before and after the vocal warm up. **All values are in dB**. (*marks significant difference, calculated with a paired t-test, $p < 0.01$)

Subject	Mean for 3-4 kHz (Before)	Standard Deviation	Mean for 3-4 kHz (After)	Standard Deviation	Difference in mean for range 3-4 kHz
1	1.8746	2.5400	2.9453	2.1978	1.0708*
2	6.5986	2.2892	7.1842	1.9451	0.5856*
3	1.8908	3.2727	2.2878	2.4017	0.3971
4	7.4969	3.2701	7.8218	2.2896	0.3249
5	4.4567	2.9868	5.6999	2.0540	1.2432*
6	5.3534	2.8222	6.1137	2.8719	0.7603*

Table 2 - Mean magnitudes in the range of 4-5 kHz of LTAS results of six subjects before and after the vocal warm up. **All values are in dB**. (*marks significant difference, calculated with a paired t-test, $p < 0.01$)

Subject	Mean for 4-5 kHz (Before)	Standard Deviation	Mean for 4-5 kHz (After)	Standard Deviation	Difference in mean for range 4-5 kHz
1	-1.8215	1.3015	4.0940	1.1984	5.9155*
2	2.8317	0.9052	5.0156	0.9938	2.1839*
3	1.5407	0.9479	2.9673	1.1408	1.4266*
4	1.8843	2.7430	3.0647	2.8041	1.1804
5	-0.2514	1.6838	1.8229	2.1136	2.0743*
6	0.7827	1.8180	1.7847	2.4227	1.0020*

3.2 Transfer Function Measurement Results

To analyze the measurement results, the average of values of formants - F1, F2 and F3 observed for transfer functions of all the vowels were compared with the average values of a large group of native German speakers provided in [9]. Table 3 shows these comparisons, where N is the number of subjects out of seven which produced interpretable results for each vowel, which proves the plausibility of the results produced with the proposed setup.

Figures 4 and 5 show some of the best and the worst transfer functions for the nine German vowels taken from the recordings of different subjects. Transfer functions in Figure 4 have smoother curves and show less interference of noise as compared to those in Figure 5, which hardly have any distinguishable formants.

Table 3 – Mean values of formants - F1, F2 and F3 of nine German vowels and reference values taken from [9], where N is the number of subjects out of seven that show interpretable results for that particular vowel. (SD – Standard Deviation, *reference only available for /ɛ/)

Vowel	N	F1	SD	$F1_{ref}$	F2	SD	$F2_{ref}$	F3	SD	$F3_{ref}$
a:	5	**810**	138.56	779	**1265.40**	107.16	1347	**3027**	120.94	2785
e:	5	**489.20**	100.90	431	**2186.40**	146.75	2241	**2850.60**	81.60	2871
i:	1	**267**	-	329	**2157**	-	2316	**2830**	-	2796
o:	4	**501.75**	111.24	438	**883**	112.15	953	**2747.30**	249.05	2835
u:	4	**411.25**	78.53	350	**931**	171.84	1048	**2626.50**	194.44	2760
ə	5	**664.60**	75.60	420	**1433.80**	140.42	1746	**2768**	114.05	2811
ɛ:	5	**701.60**	111.42	592*	**1695.60**	214.65	1944*	**2828**	197.69	2867*
ø:	6	**473.50**	73.52	434	**1589**	131.61	1646	**2475.70**	169.83	2573
y:	6	**377**	112.63	342	**1759.30**	254.30	1667	**2582.30**	91.51	2585

While these results are an encouraging find for the further exploration of the proposed method, it was, however, not possible to identify significant and consistent changes in the transfer functions for the stage voice and the normal voice.

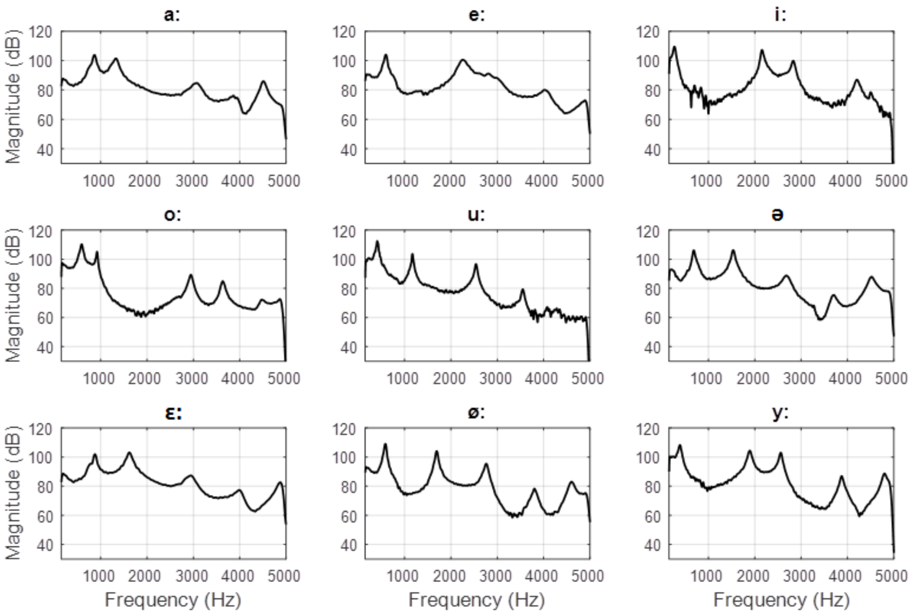

Figure 4 Best case measured transfer functions for nine German Vowels chosen from different subjects. The measurements were made with the software Measure transfer Function.

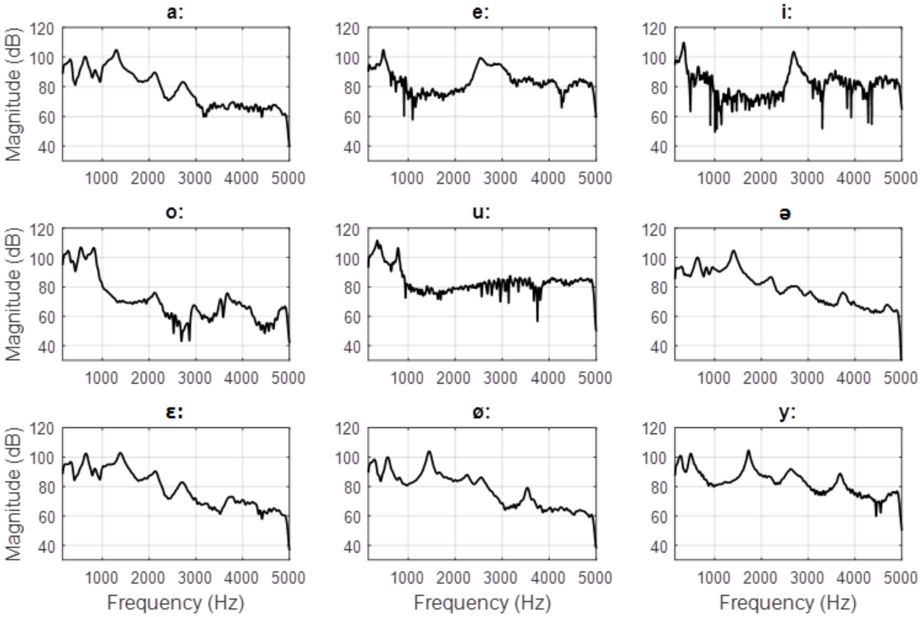

Figure 5 Worst case measured transfer functions for nine German Vowels chosen from different subjects. The measurements were made with the software Measure transfer Function.

4 Conclusion and Outlook

In the LTAS results of the female subjects, there is a considerable difference in magnitude after the vocal warm up. The effect of vocal exercises on the LTAS results has been shown by Leino *et al.* in [1] through a 30 minute vocal warm-up. The present results show the same effect caused by the vocal exercises even though they were performed for a shorter duration and a different set of exercises. The 3-4 kHz range has been established for a male Actor's Formant [1] but the female Actor's Formant may not necessarily lie in the same region. All of the LTAS presented in this study showed a larger (and more often significant across the subjects) increase in energy in the region of 4-5 kHz than between 3-4 kHz when comparing pre- and post-warm up (i.e., non-stage and stage voice) recordings. The results also support Leino's findings (publically inaccessible) reported by proxy in [4], who found a 4.3 kHz peak in actresses' voices.

To put the results into context, it is however important to consider the effect of varying loudness in speech on the LTAS of a female voice as shown in [10], because the increase in magnitude between 4-5 kHz may be due to loudness variation. The conclusions from this study therefore need to be validated by another experiment where loudness is controlled. While generally producing encouraging results, the proposed experimental setup for the measurement of vocal tract transfer functions did not help in the identification of an Actor's Formant. A major problem with this procedure was the lack of auditory feedback during the measurements because the subjects were required to stop the phonation for the sweep duration. Therefore, a significant improvement of the system would be to allow the phonation of sounds and the simultaneous measurement of the transfer function. Further research into possible algorithms to separate the filtered sweep from the subject's own speech is necessary.

It was difficult to record a usable /i:/ transfer function which is most likely due to the weak signal strength of the output signal at the lips as a result of a small lip opening. Further improvements need to be made so that the output signal has sufficient strength provided that the cross-talk is suppressed.

It was also evident in the measured transfer functions that the formants F4 and F5 were only distinguishable for a few subjects and the higher frequency regions were generally less clear. This may be alleviated by changing the excitation source from a membrane loudspeaker to a different electro-mechanical transducer, e.g., a shaker, because the transcutaneous coupling has a low-pass effect that may be more difficult to overcome for the speaker than for the shaker.

5 Acknowledgments

The experiment was performed at a studio at the Department of Speech Science and Phonetics, Martin-Luther-Universität, Halle-Wittenberg. The authors thank the subjects for their participation and support. The assistance and countless efforts of Susanne Drechsel in organizing and helping with the experiment are greatly appreciated. The research stay of the main author of this paper at the Technische Universität Dresden was funded by the Association of Friends and Sponsors of TU Dresden.

References

[1] LEINO, T., A.-M. LAUKKANEN and V. RADOLF: *Formation of the actor's/speaker's formant: a study applying spectrum analysis and computer modeling.* In: *Journal of Voice* 25.2 (2011), pp. 150-158.

[2] LÖFQVIST, A. and B. MANDERSSON: *Long-time average spectrum of speech and voice analysis*. In: *Folia Phoniatrica et Logopaedica* 39.5 (1987), pp. 221-229.

[3] CLEVELAND, T. F., J. SUNDBERG and R. E. STONE: *Long-term-average spectrum characteristics of country singers during speaking and singing*. *Journal of Voice* 15.1 (2001), pp. 54-60.

[4] MASTER, S., N. G. DE BIASE and S. MADUREIRA: *What About the "Actor's Formant" in Actresses' Voices?*. In: *Journal of Voice* 26.3 (2012), pp. e117-e122.

[5] BADIN, P.: *Fricative consonants-acoustic and X-ray measurements.* In: *Journal of phonetics*, 19.3-4 (1991), pp. 397-408.

[6] PHAM THI, N. Y.: *Caractérisation acoustique du conduit vocal: Fonctions de transfert acoustique et sources de bruit (Etude des voyelles chuchotées et des consonnes fricatives non voisées. Doctoral dissertation* (1995).

[7] LAUKKANEN, A. M., J. HORÁČEK, P. KRUPA and J. G. ŠVEC: *The effect of phonation into a straw on the vocal tract adjustments and formant frequencies. A preliminary MRI study on a single subject completed with acoustic results.* In: *Biomedical Signal Processing and Control* 7.1 (2012), pp. 50-57.

[8] TITZE, I.: *Introducing a video for using straw phonation*. In: *Journal of Singing* 66.5 (2010): p. 559.

[9] PÄTZOLD, M. and A. P. SIMPSON: *Acoustic analysis of German vowels in the Kiel Corpus of Read Speech. The Kiel Corpus of Read/Spontaneous Speech Acoustic data base, processing tools and analysis results.* In: *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel (AIPUK)*, 32 (1997) pp. 215-47.

[10] SUNDBERG, J. and M. NORDENBERG: *Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech.* In: *The Journal of the Acoustical Society of America* 120.1 (2006), pp. 453-457.