# INVESTIGATING LARYNX HEIGHT WITH AN ARTICULATORY SYNTHESIZER

*Eva Lasarcyk*

Institute of Phonetics, Saarland University, Germany
evaly@coli.uni-saarland.de

## ABSTRACT

In this paper, we present a comparative study of natural and synthetic speech samples which vary in larynx height. The acoustics of isolated vowels was analyzed with respect to formant frequency changes and changes in voice quality.

The synthetic stimuli show the same characteristics as the natural stimuli when special attention is paid to the synthetic excitation quality.

One issue addressed behind this study is how the naturalness of speech synthesis can be improved by manipulating voice quality. Another issue is to find out how well the articulatory speech synthesizer used matches the real speech production process.

**Keywords:** articulatory synthesis, naturalness, larynx height, formant frequencies, voice quality

## 1. INTRODUCTION

In this study, we investigate the acoustic correlates of larynx height in natural speech production in three larynx settings: raised larynx, neutral, and lowered. Then, we compare them to articulatorily synthesized versions. This is to find out about the changes in the acoustics of (human) larynx height in terms of synthesis parameters such as breathiness.

The main actor for vertical changes of larynx height is the hyoid bone [4, p. 24ff.]. This bone is an interaction point of several different muscular systems in speech. This leads us to the assumption, that merely changing larynx height in the articulatory synthesizer will not be sufficient to achieve the same acoustic outcome as a natural speaker. We will also apply changes to the voice quality in the synthesizer and compare the results.

The approach used here is to change the degree of breathiness with larynx height. Observations suggest that lax voice is often accompanied by a slightly breathy voice: When the infrahyoid muscles pull down the larynx, the suprahyoid muscles should relax – this relaxation is also found in the production of breathy voice [4, p. 31].

Appropriate voice quality is a core factor for natural and adequate sounding synthetic speech. Along this line, our goal is to find out how well an articulatory synthesizer does in simulating the desired properties of voice quality associated with larynx height. Variation of larynx height is e.g. used in Asian languages as part of register [3] and is probably also involved in smiled speech.

## 2. ACOUSTICS OF LARYNX HEIGHT

To illustrate the changes that take place when a speaker varies larynx height, we analyzed a small sample of natural speech. The speaker was asked to produce isolated vowels and to focus on larynx height control. The experimenter checked this visually and auditorily. Other vocal tract features for segmental articulation were kept as constant as possible. In spite of these intentions for constant articulation, it is assumed that some features in the speech production process do change as well.

### 2.1. Changes of formant frequencies

Varying larynx height changes the overall length of the vocal tract. This will directly affect the formants of the vowels produced [4]. Compared to speech in a neutral setting, the formants should be

- Higher in a raised larynx setting, and
- Lower in a lowered larynx setting.

### 2.2. Changes of voice quality

Apart from the straightforward influence on formant frequencies, natural speech is influenced with respect to voice quality. The substantial vertical shift of the larynx affects the mode of vibration of the vocal folds [4][8][9]. It is assumed that

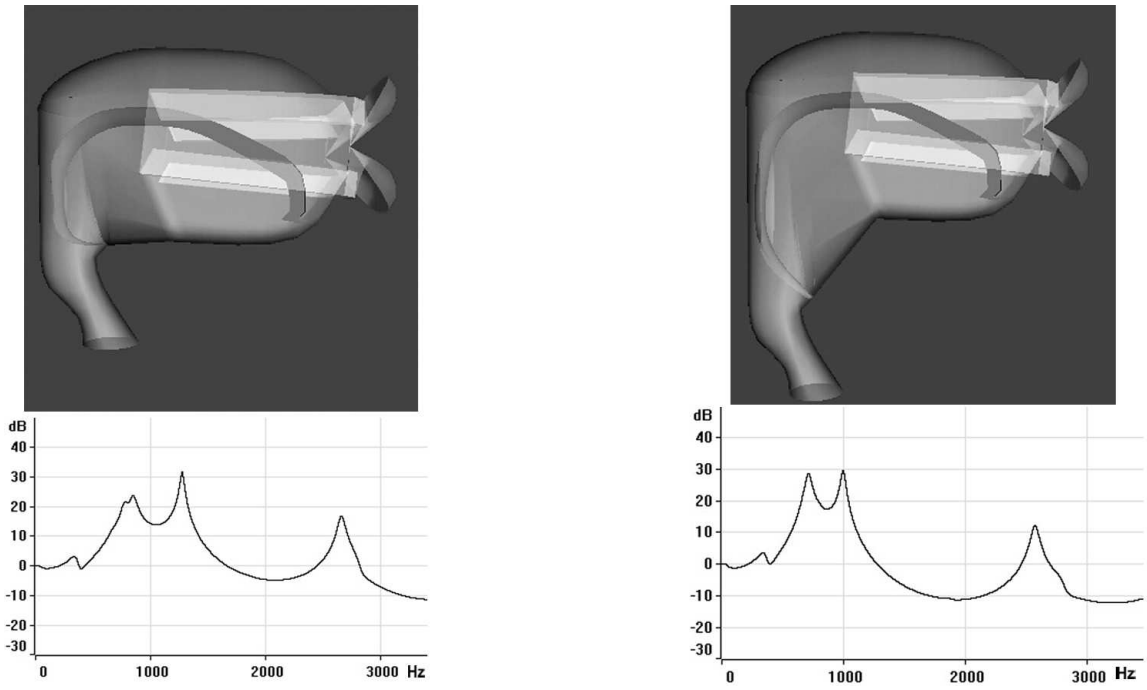
- Raised larynx speech sounds tenser, and
- Lowered larynx speech sounds laxer

compared to the speech in the neutral setting (modal voice) when the larynx is between the two extreme positions. Acoustically, this can be confirmed by measurements used to distinguish voice qualities (cf. section 4).

## 3. THE ARTICULATORY SYNTHESIZER

For the generation of the synthetic stimuli, the articulatory synthesizer in [1] was used. It is based on a 3D articulatory model of the vocal tract

**Figure 1:** Vocal tract configuration and transfer function for [a:] in raised larynx setting (left) and in lowered larynx setting (right). Note the differences in the pharynx and in the formant frequencies.




combined with a partially novel aerodynamic-acoustic simulation [1, p. 101]. Articulation is controlled by "gestures". On a gestural score, the timing and magnitude of the articulatory gestures are specified. The vocalic and consonantal gestures are associated with particular configurations of the vocal tract (articulatory goals).

The vocal tract configuration can be changed by altering the parameters of certain control points in the articulators. The software version used here supplies German phoneme settings; it is adaptable though to any other language.

Fig. 1 illustrates the changes made to the configuration of the vowel [a:] with respect to larynx height in order to synthesize the vowel stimuli. The main difference to be noted is located in the pharynx. The graph below the vocal tract model depicts the transfer function of the corresponding vocal tract configuration, illustrating that raised larynx raises the formants and lowered larynx lowers them.

**Table 1:** Number of natural speech vowel tokens at each larynx height taken for analysis.

| | [a:] | [i:] | [u:] |
|---|---|---|---|
| Raised | 2 | 5 | 5 |
| Neutral | 2 | 3 | 3 |
| Lowered | 1 | 2 | 3 |

By specifying glottal gestures, different degrees of breathiness can be added to the signal. This makes it possible to generate the breathy vowels needed for this study.

## 4. SPEECH MATERIAL AND ANALYSIS

### 4.1. Recordings

#### *4.1.1. Natural speech*

As stated in section 2, the natural speech examples were recorded by one adult male speaker. The isolated vowels [a:], [i:], [u:] were part of a corpus comprising vowels, a word list and a number of sentences. The corpus was produced with a raised larynx, a lowered larynx, and with a normal larynx position. Isolated vowel realizations for each larynx height were selected from those parts of the recordings where experimenter and speaker agreed that the larynx setting had been optimal (see Table 1).

#### *4.1.2. Synthetic speech*

The same vowels were synthesized with the articulatory speech synthesizer. The larynx height was altered as illustrated in Fig. 1. To simulate the different excitation qualities that accompany

changes in larynx height, the degree of *ad*duction for the voiced intervals was adjusted to give differing degrees of "breathiness": 0 for tense voice, 5 for modal voice, and 10 for lax voice (on a system scale of 0-100, where 100 equals complete *ab*duction; cf. image files 1 and 2). Thus 3 times 3 stimuli of each sample were created: Each larynx height (lowered, neutral, raised) combined with each degree of breathiness (none, slight, moderate).

### 4.2. Data analysis

#### *4.2.1. Raw data points*

Both kinds of speech data, natural and synthetic, were subjected to the same kinds of analyses.

A stationary part (regarding F1, F2, and F3) of each of the vowels was analyzed with Praat [6] with respect to the following aspects:

- Formant analysis of the first three formants (F1, F2, F3)
- Visual inspection of the amplitude spectrum (FFT) of a spectrum slice covering that same stationary part of the vowel. In this way the dB values of the first two harmonics (H1, H2) and the first three formants (A1, A2, A3) were extracted. No smoothing was applied to the spectrum slice and the amplitudes of the formants were taken from the value of the closest harmonic.

#### *4.2.2. Data processing*

The following values were first calculated and averaged for each vowel and then averaged over all vowels. The values of the formant frequencies were taken directly for analysis of formant changes with larynx height.

The amplitude values were used to calculate the following voice quality measurements [2].

- Amplitude differences of H1 and H2 in dB (H1-H2), reflecting the open quotient, i.e. the time in which the glottis is open.
- Amplitude differences of the first harmonic and the first three formants normalized in dB per Octave: H1-A1 indicates the degree of glottal opening during the whole oscillation period of the vocal folds. H1-A2 shows how abruptly the glottis is being closed. A gradual cutoff leads to a greater loss of energy in the higher frequencies. H1-A3 complements the latter in that it characterizes the velocity at which the airflow is cut off (Skewness and Rate of Closure in [2].)

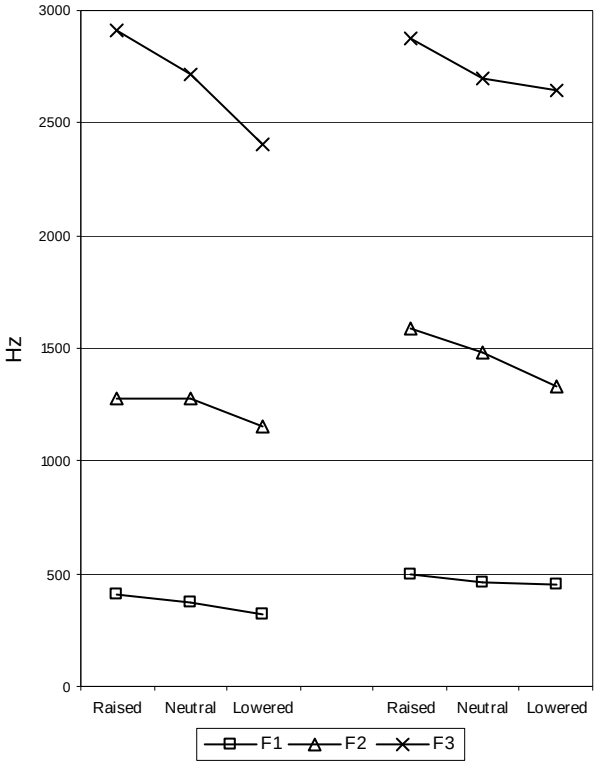



**Figure 2:** Formant frequencies of vowel stimuli over three larynx settings (Raised, Neutral, Lowered): Natural speech (left), synthetic speech (right).

All in all, higher amplitude differences indicate a more lax voice quality and can also be associated with a lowered larynx setting [2].

## 5. RESULTS

### 5.1. Formant frequencies

As expected, the formant frequencies in the natural speech samples get lower as larynx height decreases. The same holds for the synthetic vowel frequencies (see Fig. 2).

### 5.2. Voice quality measurements

In the samples of natural speech, all voice quality measurements fall (i.e. the spectrum is flatter) when larynx height is increased. This indicates that the elevation of the larynx is indeed accompanied by an increase of phonatory tension (cf. section 2). The same tendency can be observed in the synthetic speech samples (see Fig. 3), where the H1-H2, H1-A1, H1-A2, H1-A3 intensity differences are shown to be similar for the natural productions and the synthetic stimuli combining raised larynx with 0-breathiness, normal height with 5-breathiness and lowered larynx with 10-breathiness, as motivated in section 1.

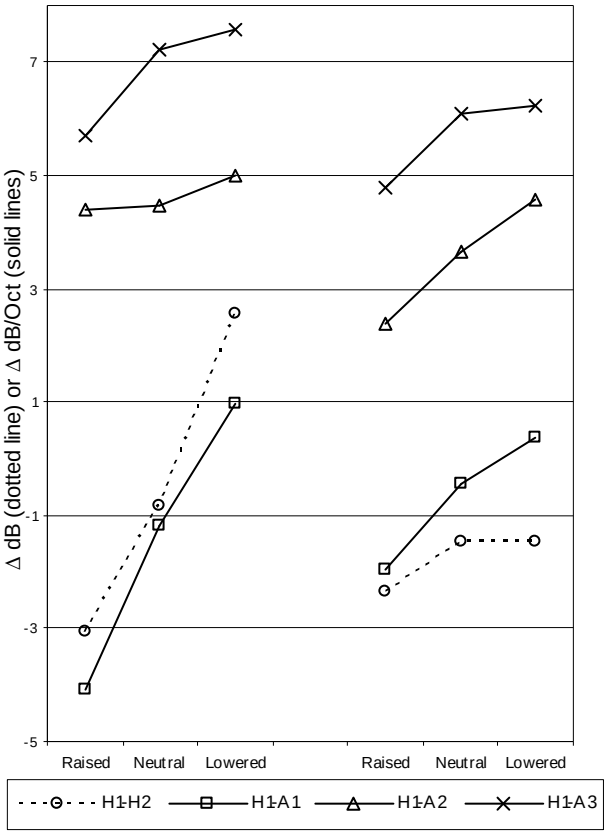



**Figure 3:** Voice quality measurements of vowel stimuli over three larynx settings (Raised, Neutral, Lowered). Natural speech (left), synthetic speech (right). In synthetic speech, breathiness is varied (cf. section 5.2 and 4.1.2)

Combining the three larynx heights with a single excitation quality setting results in spectral tilt values (H1-A1..3) that only change slightly with the changing formant values and do not reflect at all the changes observed in natural speech (cf. image file 3).

This implies that it is not sufficient to manipulate larynx height alone in the model. You have to change the excitation characteristics in order to get voice quality characteristics comparable to natural speech.

## 6. DISCUSSION

As a role model for synthetic vowels, a set of natural vowels differing in larynx height were analyzed with respect to formant frequency changes and voice quality changes over larynx height. The findings in the natural vowels could then be reproduced in synthetic vowels generated with an articulatory speech synthesizer. In both cases, a lowering of formant frequencies can be observed as well as a change in voice quality from tense to lax.

Firstly, we learn from the synthetic reproduction how larynx height affects human speech sound production by making explicit synthetically what only co-occurs in natural speech production processes. Since the synthesizer uses a geometrical 3D model, every change in the vocal tract configuration has to be made explicitly. The interdependencies of articulators and glottis excitation characteristics that we find in humans have to be taken care of individually in the synthetic vocal tract. In this way, we explicitly added varying degrees of breathiness in different larynx heights to resemble the tenser or laxer voice quality found in the natural speech samples.

Secondly, this shows that the articulatory synthesizer [1] is able to reproduce speech with sufficient detail to be acoustically comparable to natural speech stimuli, in particular with regard to voice quality issues. This is a fact worth noting when it comes to the issue of addressing and improving naturalness in speech synthesis systems [7].

## 7. REFERENCES

[1] Birkholz, P. 2005. *3D-Artikulatorische Sprachsynthese.* University of Rostock (PhD thesis).

[2] Claßen, K., Dogil, G., Jessen, M., Marasek, K., Wokurek, W. 1998. Stimmqualiät und Wortbetonung im Deutschen.
*Linguistische Berichte* 174, 202-246.

[3] Edmondson, J.A., Esling, J., Harris, J.G., Shaoni, L., Ziwo, L. 2000. The aryepiglottic folds and voice quality in the Yi and Bai language: laryngoscopic case studies. *Mon Khmer Studies* 31, 83-100.

[4] Laver, J. 1980. *The phonetic description of voice quality.* London: CUP.

[5] Neppert, J. 1999. *Elemente einer akustischen Phonetik.* Hamburg: Buske.

[6] Praat: doing phonetics by computer. http://www.praat.org/ visited 5-Mar-07

[7] Shadle, C. H., Damper, R. I. 2002. Prospects for articulatory synthesis: A position paper. *Proceedings of 4th ISCA Workshop on Speech Synthesis,* August/September 2001, Pitlochry, Scotland, 121-126.

[8] Strik, H., Boves, L. 1992. Control of fundamental frequency, intensity and voice quality in speech. *Journal of Phonetics* 20, 15-25.

[9] Sundberg, J., Askenfelt, A. 1981. Larynx height and voice source. A relationship? *Dept. for Speech, Music and Hearing. Quarterly Progress and Status Report* 22, 23-36.