INTRODUCTION
The human voice is a fundamental element for social interaction and self-expression, and the combination of well-defined vowels with consonants is the base for any human language. Successful voice development relies on learning new—or relearning previous—vocal motor behaviors which result in optimized output qualities (eg, loudness, pitch, voice quality, respiration, efficiency). Voice production, especially during singing, is a complex process where different aspects such as the stance, breathing, glottal activity, and shaping of the vocal tract are simultaneously addressed.

The sound of a healthy voice can easily be recognized; however, it is desirable to rely on clearly defined objective measurement procedures to monitor the state of the voice and its development for each of the variables independently. There is a large variety of analytical methods that can be used for respiratory analysis or posture, on the one hand, and of the voice during phonation, on the other hand. In regard to feedback procedures on phonation, microphone measurements are widely used to improve pitch and loudness of the singer or speaker with short response time.

Among other elements essential for voice production, the highly flexible vocal tract is the key element for vowel production. An adjustment of its resonances (formants) is possible by means of the articulatory elements (ie, lips, tongue, jaw, and larynx).

The present work discusses a possible fast biofeedback mechanism for an adjustment of the vocal tract shape toward specific frequencies within the range of the first and second formants. For this purpose, an external sound field in front of the mouth is offered and the participant is asked to actively adapt the articulatory elements of the vocal tract to yield a maximum amplification without phonation (ie, without oscillation of the vocal folds).

In the literature, there are several approaches to measure acoustic data of the vocal tract. Fujimura and Lindquist excited the vocal tract via the neck obtaining the information about the transfer function. The reverse measurement is also possible using a three-dimensional model of the vocal tract derived from MRT studies. In this case, the plastic model is loaded by an external sound source at distance and a microphone located at the position of the glottis.

Rothenberg used a tube close to the mouth driven by a membrane to stimulate vibrations across the vocal tract at the vocal folds. Highly accurate impedance data in a broad wavelength spectrum could be realized via an impedance tube even during phonation. Epps et al introduced a method to measure the load of the vocal tract with an acoustical duct close to the mouth and a microphone at a similar position. In this arrangement, the impedance of the resonator (vocal tract) can be separated from the output impedance of the sound source (loudspeaker) by introducing a high resistive load within the acoustic duct, limiting the effects of tube resonances.

In our study, a sound source close to the mouth and similar to the ones used in mobile phones interacts directly with the vocal tract of the participant at specific frequencies. The sum of the acoustic signals of the external sound source and the amplification due to the resonance of the vocal tract are used as biofeedback signal. Furthermore, the effect of the amplification on the acoustic properties of the vocal tract is discussed.

EXPERIMENTAL PROCEDURE

Instrumentation
Acoustical measurements have been conducted within an audio booth of the phonetics department of the TU Munich or the Fraunhofer-Institute of Integrated Circuits in Erlangen, the latter with a residual sound level of less than 20 dB and spectrally linear reverberation. A Schoeps (Germany) MK2 microphone with a preamplifier CMC 6 was used,
which was confirmed by a GRAS 46AF (G.R.A.S Ltd. UK) microphone connected via a preamplifier Nexus 2690 (Bruel&Kjaer, Denmark) to a Steinberg UR242 (Yamaha Inc., Japan) as an external sound card. The distance between the microphone and the vocal source was 20 cm. All recordings were performed at 48,000 Hz/16 bit using the software Audacity (D. Mazzoni, Mountain View, CA, USA).

Spectrograms and Fast Fourier Transform (FFT) spectra were obtained using the WaveSurfer (Sjolander and Beskow, Stockholm, Sweden) Version 1.8.8.p4 software. The FFT analysis during phonation was performed at wavelength steps of 24 Hz (2048 FFT points), spectrogram of the exhaled air and under phonation was analyzed with an analysis bandwidth of 4 Hz.

The quality factor may also be interpreted as energy stored in the resonating wave. The resonance in respect to frequency will be met if the frequency of the driver meets the resonance frequency $f_0$. Furthermore, damping will broaden the resonance in respect to $\omega$ and reduce the amplitude. The amplification (resonance) of the resonator (i.e., the vocal tract) is thus a synonym for low bandwidth ($B_w$) or high quality factor ($Q$) at the given frequency. In this case, the maximum amplification $A$ will be limited by the loss factor $\alpha$ by

$$A_{\text{max}} \sim \frac{1}{\alpha} \sim \frac{f_0}{B_w} = Q. \quad (1)$$

The quality factor may also be interpreted as energy stored divided by the energy dissipated. It is obvious that high
amplitudes will be obtained at the resonance frequency and minimized loss factor.

The phase lag $\varphi$ between acoustical driver and the amplified sound of the resonator is determined by

$$\tan \varphi = \frac{2\omega_0}{\omega_0^2 - \omega^2},$$

with $\omega$ being the angular frequency $2\pi f$ of the driver. Thus, a phase shift of $-\pi/2$ is observed if frequencies of driver and resonator are equal. The phase shift is zero either with no coupling between resonator and driver or the resonator is tuned to a significant higher frequency. A phase shift of $-\pi$ will be found at significantly lower resonance frequencies of the resonator compared with the driver frequency.

The vocal tract is regarded as an open-closed tube which resonance frequency is tuned to an externally offered sinusoidal signal. The resonance conditions of an open-closed tube are determined by the radiation impedance at the tube opening, which can be described with the assumption of $\alpha L \ll 1$, $\frac{\alpha}{k} \ll 1$ by

$$Z_{m0} = \frac{\alpha L - i \cos kL \sin kL}{\sin kL + (\alpha L)^2 \cos^2 kL},$$

where $Z_{m0}$ is the radiation impedance at the tube opening, $L$ is the length of the closed tube including an end correction, $k = 2\pi / \lambda$, and $\alpha$ is the loss coefficient.

The complex and the real parts of the impedance at the tube opening are high at $kL = n \pi$ and the radiation is small (antiresonances). At the resonance frequency, the imaginary part of the impedance will be zero at $kL = (2n + 1/2)\pi$, with a pressure node at the open end and a pressure antinode at the closed end. In relation to the particle velocity, an antinode is formed at the open and a node at the closed end. With the imaginary part being zero at resonance, the impedance is reduced to its real part and the amplitude is determined by the loss factor $\alpha$. The real part of Eq. (3) at resonance conditions may thus be related to the sum of different losses (ie, reflection losses, wall interactions, radiation, and viscous or thermal absorption).

As stated previously, a driver with its radiation impedance $Z_{m0}$ at the tube opening can excite the acoustic system toward oscillation. In order not to alter the resonance conditions of the vocal tract, the impedance of the driver should be low, which will hold true for acoustic coupling via air and with a high particle velocity at the interface between the driver and the open mouth.

Coupling will be most effective if the driving sound source is located within the end correction term of a tube opening, that is, within a distance of about $0.8^*r$, with $r$ being the radius of the mouth opening. In the present study, this was realized by using a speaker close to the mouth. In addition, resonances of any waveguide can be neglected if the distance of the oscillating membrane of the speaker to the open mouth is chosen to be within the lumped region at the wavelengths used.

**RESULTS**

Figure 2 shows the changes in the sound level during the adjustment of the vocal tract shape toward frequencies of 700, 900, and 1300 Hz, respectively. Alterations within the articulatory shape of the vocal tract result in strong changes of the overall sound level, and the participant needed some time to find the position of the highest amplification. The procedure of adjusting the vocal tract shape, as shown in Figure 2, was repeated until a stable position could be reproduced (adjustment period).

Using a lock-in amplifier, the phase and amplitude of the resulting sound wave can be detected simultaneously. Apart from an increase in amplitude, the phase of the signal used as biofeedback is shifted as shown in Figure 3.

The sound pressure amplitude is clearly linked to the phase shift between the driving source and the amplification due to the resonance of the vocal tract. In all cases, the highest amplitude may be found at a phase shift of the resulting wave at about $-\pi/2$. In the case of the resonance amplification at 1300 Hz, the formation of antiresonance at the same phase shift and minimum of sound pressure is observed as well. In the following table,
typical values of resonance amplification are listed (anti-
resonances are omitted).

<table>
<thead>
<tr>
<th></th>
<th>700 Hz</th>
<th>900 Hz</th>
<th>1300 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amplification at −90° phase shift (dB)</td>
<td>8−16</td>
<td>7−15</td>
<td>6−17</td>
</tr>
<tr>
<td>Phase shift at highest amplitude (°)</td>
<td>−(80−100)</td>
<td>−(80−100)</td>
<td>−(80−100)</td>
</tr>
</tbody>
</table>

It is evident that an adjustment of the vocal tract toward one frequency will affect the acoustical environment of the voice production system as a whole. To observe changes in the vocal tract without using the vocal folds, the resonance spectrum of an exhaled air might be a suitable indicator. Figure 4 shows sound pressure and spectrogram of the exhaled air after adaptation of the vocal tract. Note that no vowel has been intentionally formed and the position of the highest resonance intensity was asked to be maintained.

After the adjustment period, a stable position could be sustained for several seconds and resonances are observed at the adjustment frequencies which could be reproduced within 10% of the frequency applied. The maxima of what might be attributed to the first formant shifts from around 700 to about 450 Hz as the adjustment frequencies increase. In addition, a higher resonance is observed to shift from around 2400 Hz (adjustment frequency: 700 Hz) to lower values.

Figure 5 shows the sound pressure and spectrogram during phonation after the adjustment periods. Although there is no intentional articulation of any vowel, the sound after adjustment toward 700 Hz is similar to an “o,” whereas the sound after 900 Hz has an appearance of an “a” with a tendency toward “ae” after 1300 Hz.

The basic features of the spectra are found similarly during phonation and exhalation, with the adjustment frequency being supported. In the spectrogram, the higher and lower regions are shifted with increasing adjustment frequencies, that is, after an adjustment of the vocal tract adjustment toward frequencies of 1300 Hz, the lower and higher resonance are found at about 500 and 2400 Hz, respectively.

Figure 6 shows the FFT spectra of the waveforms shown in Figure 5.

The resonance tuning can clearly be observed in the harmonics during phonation in the frequencies at 700, 900, and 1300 Hz. In addition, the highest harmonics of intensity are reduced from H4 to H3 and from H20 to H17 as the adjustment frequency (before phonation) is increased.

**DISCUSSION**

**Adjustment of the vocal tract shape and resonance**
The participant used the audible resonance of the vocal tract excited by an external sound source as biofeedback to adjust the articulatory elements toward a stable position without phonation.

In the model of a forced oscillation (Figure 1), the external speaker is considered as driver and the vocal tract as resonator. The most intense coupling will demand that (1) the resonance frequency is met and (2) the bandwidth of the resonator is small, the latter being related to a small loss coefficient. In the experimental procedure used, the driving frequency is kept constant, whereas the vocal tract of the participant (resonator) is adjusted to meet resonance conditions. As the participant was asked to enhance the amplitude of the external sound source (Figure 2), the most apt condition can be found within the first 10 seconds. In the case of the adjustment toward 900 Hz, a reduction of the overall amplitude at 2 seconds is observed. In this case, the vocal tract seems to have been tuned shortly to an antiresonance at 900 Hz.
With the mouth closed, only the amplitude of the driver is measured. In the case of the opened mouth, coupling occurs via air and the additional sound from the oscillator adds to the one of the driver. As shown in Figure 3, a phase shift is observed if the vocal tract is shaped toward resonance at the given frequency. A maximum amplitude is found at a phase lag of $\pi/2 \cdot \frac{1}{\alpha}$ (90°) between the loudspeaker and the resonator. From Eqs. (2) and (3), a phase lag other than $-\pi/2$ may be directly linked to the difference between the driving frequency of the speaker and the resonance frequency of the vocal tract.

As the lock-in amplifier used in Figure 3 records the sum of the acoustic signals—the one of the driver and the resonant signal of the vocal tract simultaneously—the phases of the waves add as well. At maximum amplification between 16 and 18 dB in relation to the acoustic signal of driving source alone (closed mouth), the observed shift is reduced by only $10^\circ$ to $-80^\circ$ but will be higher at low amplification. This would be in accordance with the values in Figure 3 at lower intensities between $-\pi/2$ and $-\pi$. However, the phase lag of less than $-90^\circ$ may also be influenced by the presence of other resonances (ie, formants) in the vicinity of the adjustment frequencies.

However, the amplitude depends not only on the resonance matching but also on the loss coefficient $\alpha$. Thus, besides changing the resonance condition of the vocal tract to meet the adjustment frequency, high amplitudes at a phase lag of $-\pi/2$ are interpreted as a synonym for a low loss coefficient (ie, a small bandwidth, a high quality factor) (Eq. 1). It is evident that a low loss coefficient will be helpful for facile voice production, which may be supported by using the proposed biofeedback.

From the experiments, it can be deduced that the loss coefficient of the vocal tract may be reduced by vocal tract adjustment without phonation. However, the measured results do not specify the origin of the effects, leading to reduced damping. In a tube-like system, viscous and thermal losses are usually low. Main losses occur via contact of the oscillation air with the mucous membranes, the radiation of sound, the shape of the vocal tract, and the reflectivity at the closed end. Although the tissue is regarded as acoustically reflective at frequencies higher than 500 Hz, the radiated power will increase with frequency. Given a constant wavelength of the signal and only small variations of the mouth opening, radiation losses may be considered as stable. Thus, it is reasonable to assume that losses due to wall interactions or radiation during resonance may only be addressed within a narrow region by the adjustment. Consequently, the shape of the vocal tract and a high reflectivity especially at the pressure nodes or antinodes will be discussed in more detail.

Equation (3) demands a pressure node at the mouth and a high impedance (reflection) condition at the closed end for the evolution of stable “standing” waves (ie, resonance). Although a pressure node close to the mouth can be assumed to prevail as a resonance condition and the location where the sound waves are reflected, the acoustical boundary conditions inside the vocal tract are less obvious. The basic physiological structure of the vocal tract may give a hint as where to find possible (and tunable) zones for high amplification of the induced resonances. In the literature, there is strong evidence that during phonation, as well as during whispered speech, an antipressure node is located at the glottis. However, there might be more flexibility to meet this condition without phonation. As all experiments took place with an unrestricted airflow from the trachea through the lips, the vocal folds may not be regarded as a high impedance barrier. The congruence of the resulting spectra after resonance adjustment by exhaling air and during phonation at frequencies of 700–1300 Hz, however,
makes it reasonable to look for anatomic structures that will be able to represent a tunable high impedance reflection barrier not far from the high impedance resonance condition present during phonation. Hanna et al. studied the acoustic properties during phonation in comparison with the glottis closed and found little changes between the two positions in regard to the transfer function of the vocal tract. The epilarynx tube with its small area function compared with the pharynx has been discussed as a possible barrier and high impedance source for a conversion of the airflow through the glottis toward high pressure at the bottom of the vocal tract. In this context, the high impedance at the transition of the pharynx to the epilarynx tube may act as a restriction with high reflectivity and thus the proposed pressure antinode. Furthermore, if the high impedance necessary for resonance condition is found in this region, an adjustment of the acoustical behavior toward external frequencies should have a positive impact on the coupling of the vibration of the vocal folds with the vocal tract as well. In this context, it should be noted that during tube phonation, a phase lag between the driver (vocal folds) and the vocal tract of $-\pi/2$ at maximum coupling has been discussed.

The proposed driver-resonator-model related to Eqs. (1) and (2) is not linked to any specific geometry of the vocal tract. Apart from the previously discussed tube model, where adjustment of the vocal tract shape is directly related to the length of the resonator, low bandwidth resonance may also be achieved in Helmholtz resonators. In a Helmholtz resonator, the volume in a small neck represents the mass of the vibrating system and is connected to a volume...
of spherical size, acting as a spring. Also, in this model, a pressure node is present at the mouth opening. The transition between a tube model and a Helmholtz resonator within a model pipe system with two different diameters occurs gradually. This more "Helmholtz-like" position may be realized with a large area function at the pharynx compared with the area function of the mouth.

Vocal tract adjustment and formants

Human vowels are identified by the position of the first and second formants, which can be interpreted as vocal tract resonances (Figure 1). As the concept of format formation is rather related to the shape of the vocal tract than to vowel recognition, we call the observed increases in intensity "formants" in the following discussion.

During the adjustment of the vocal tract, only one frequency is actively amplified. However, other resonances are affected during the positioning, also, as can be seen in the spectra of the exhaled voice and during phonation. Note that no intentional change in the vocal tract position was made and the adjustment took place at normal breathing.

Studies showed rising formants in the whispered voice compared with phonation, and the changes were explained by a reduction in the diameter of the vocal tract close to the glottis. In Figure 3, the transfer function at open glottis condition during exhaled air is close to the position during adjustment of the vocal tract shape. Exhaled air seems to preserve the vocal tract resonances, whereas whispering may be regarded as being related to intentional changes within a separate communication mode. It is surprising, however, that the frequencies used for adjustment find their representation during phonation as well. Obviously, there must be changes from the open vocal folds toward a position supporting phonation with the folds partially closed. However, we can regard resonance as the condition where sound waves reflected from the closed and the open end result in a constructive interference (Figure 1, "standing" waves). In this concept, the location of excitation of the resonator itself will not change the resonance frequencies as long as the excitation occurs compatible with the resonance conditions. With the excitation of the vocal tract during the adjustment phase with high particle velocity at the open mouth, this condition should be fulfilled. The proposed mechanism should be supported by the high quality factor of the vocal tract induced during the adjustment phase.

After an adjustment frequency of 700 Hz, the first and the second formants are found close to each other (phonated sound similar to an "o"). As the adjustment frequency is increased, the second formant rises in parallel frequency, whereas the frequency of the first formant is lowered. The increases in intensity of the second formant are most pronounced. As the adjustment frequencies of 900 and 1300 Hz address mainly the second formant, a higher impact can be expected.

Furthermore, the intensity of the formant region associated with the singer's formant clusters is increased with increased adjustment frequency (Figure 6). Fant pointed out that small frequency differences between the second formant and the singer's formant cluster may be favorable to higher intensities in the area of 2.5 kHz.

As small formant bandwidths are useful for better intelligibility or articulation, the tuning of the vocal tract toward an external sound might help to enhance vocal capabilities in spoken language as well. Furthermore, professional singers use resonance tuning strategies to increase the audibility of their voices. This mechanism is used mainly at higher \( f_0 \) range by tenors and sopranos. In the case presented, a comfortable pitch region with comparably narrow spacing of the harmonics was chosen. From these findings, it seems possible to adjust the position of a specific formant to achieve a desired amplification by the proposed biofeedback.

FIGURE 6. FFT analysis after the adjustment period \((f_0 = 131\) Hz). The analysis is a 3-second average of a selection taken at the center of the produced sounds after adjustment of the vocal tract shape to 700, 900, and 1300 Hz (Figure 5). The arrows indicate the adjustment frequency.
CONCLUSION AND OUTLOOK
In this study, the first findings on the influence of an adjustment of the vocal tract resonance conditions toward external acoustical signals via biofeedback are presented. The results are in agreement with the model of a driven resonator where the driver is represented by the external sound source and the resonator by the vocal tract. From measurements of phase and amplitude of the enhanced amplitudes, a narrowing of the formant bandwidths after adjustment is deduced. The increased amplitude of the resonance may be linked to lower damping at the frequencies considered by changes of the area function of the vocal tract. The low acoustical resistance of the vocal tract obtained via adjustment of the vocal tract shape without using the vocal folds should therefore facilitate voice production under phonation as well. Furthermore, the adjusted vocal tract preserves its acoustic features during exhalation and even phonation. Thus, highly specific amplification of single frequencies by an adjustment of the vocal tract shape may be helpful in vocal training or warm-up procedures. Moreover, the proposed adjustment may be helpful to reduce the complexity in vocal training and increase the understanding of the basic elements of voice production. Further studies need to be performed to define and evaluate procedures possibly suitable for voice development.

Acknowledgment
The authors thank Thomas Waschkies from Fraunhofer IZFP for helpful discussion. The authors are grateful for the support in experimental instrumentation by Peter Brandstaedt and Zlatko Dubovski from Fraunhofer-IBP in Stuttgart/Germany and of Karsten Buse from the Fraunhofer-IPM in Freiburg/Germany. Thanks are due to Matthias Struck for access to the audio facilities at the Fraunhofer-IIS.

REFERENCES