Influence of early recording and playing devices on voice sounds: Modification of singing voice formants

M. Kob, S. V. Amengual Garí, B.-A. Bolles, D. M. Ritter, P. Pirch
Detmold University of Music, 32756 Detmold, Germany

This manuscript is part two of a two-part publication at DAGA 2018. Whereas the first part covers the acquisition and interpretation of frequency response function (FRF) measurements of gramophone parts (horn, tonearm, soundbox) [10], this part investigates the influence of the FRFs on the formant structure of singing voice signals.

Previous Research

An investigation on the phonograph working principle has taken place within a student project at the Erich Thienhaus Institute in Detmold. A couple of cylinders were used to study the frequency response of a phonograph. A sweep signal and single tones reproduced with a loudspeaker were inserted into a conical recording horn and recorded with a wax cylinder. Then the recording was played back through an exponential playback horn and recorded with a measurement microphone. With this technique it was possible to determine the acoustic characteristics of the device. The concept is shown in Fig. 1. This set-up allows for direct comparison of the original

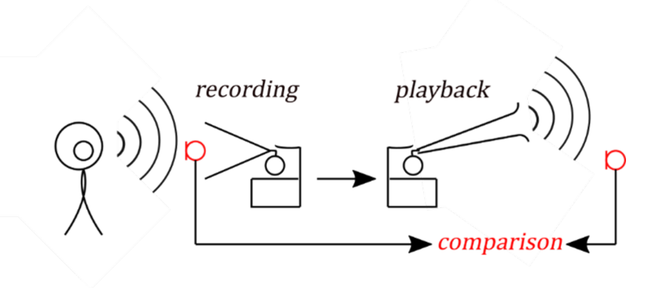

Figure 1: Comparison of voice signals using historic recording devices [6].

voice signal in front of the recording device and the reproduced recorded signal taking into account both parts: the recording and the reproduction chain. A direct comparison of the spectra using a sweep signal reproduced with a loudspeaker instead of a voice signal indicates the main properties of the historic recording devices. In Fig. 2 the spectrum of the swept sine signal reveals a strong band-pass effect of the recording, together with distortions and a significant background noise.

The first voice recording tests included distance tests, recordings of musical a cappella pieces, and selected vowels at different pitches. In figure 3 the change in voice timbre is visualized by the change in formant structure. Obviously some formants (F1, F3, F4) are not visible in the voice signal that has been recorded and reproduced

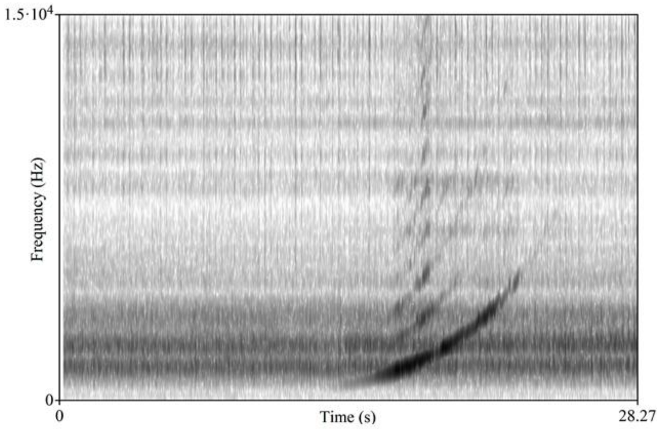

Figure 2: Swept sine recorded and played back by the phonograph [3].

with the Edison Phonograph whereas in the frequency range between F2 and F3 new formant-like maxima in the voice spectra have occurred. The details of these first investigations can be found in [6].

History of voice recording

Since the invention of electro acoustic devices such as microphones and amplifiers recording technology has not only increased the quality of voice recordings, but also extended the potential use of amplified voices and even created new singing styles such as Crooning [2] and Beatbox [7].

The mechanical and acoustical conditions of early recording devices imposed limitations on the singing performance with respect to dynamics and ornaments such as vibrato. The small range between noise floor and non-linear transmission of the signal required a well-adjusted vocal effort. It was reported from Edison that he rejected singers with too strong vibrato [5]. One goal of our research is to investigate the effect of vibrato on the recording and reproduction process.

Transfer path analysis

The key difference between a phonograph as a recording device and a microphone is that an output from a contemporary microphone is a small voltage variation representing the sound pressure change, whereas the membrane of the phonograph has to move the stylus and engrave the trace in a stiff wax medium using the actual energy of the sound wave.

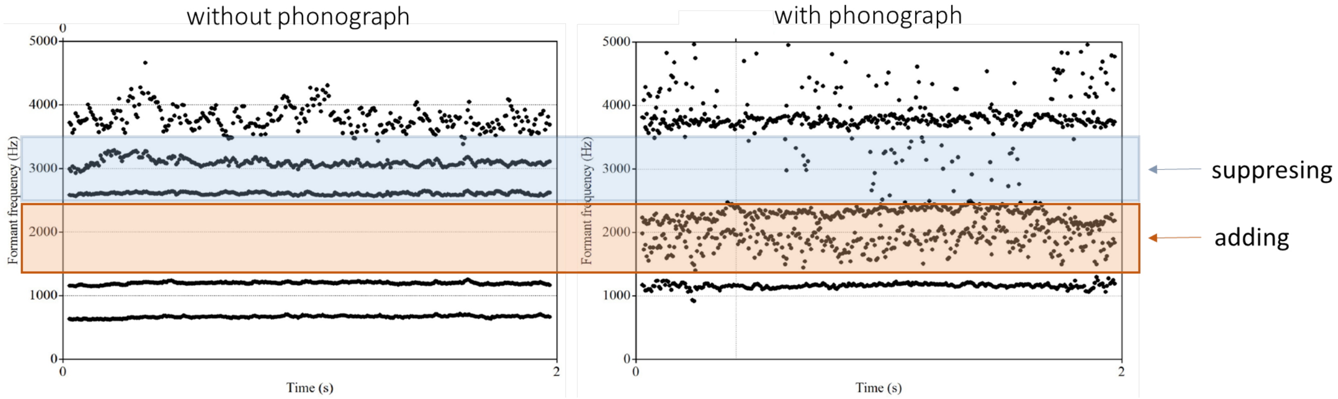

Figure 3: Differences in formant structure between modern (left) and Phonograph (right) recording

In a conventional microphone, the aim is to place the resonances of the membrane outside the audible range to achieve an even p-V transfer function. In phonographs and gramophones, the use of horns with their resonances and limited transfer capabilities is the only way to collect enough energy to engrave a wax record. Also, the sound box acts as a resonator and band pass filter. At low frequencies, there is not enough acoustic energy to move the recording stylus, and at high frequencies, the viscosity of the wax is an obstacle. More details of these transfer characteristics are given in [10].

Formant changes

The effects of the phonograph on the sung voice were found both in spectral and time domain.

Spectral domain: The frequency range of the device is from about 500 Hz to 4500 Hz. In a phonograph record, the vocal formants are located in this range, and the fundamentals below 500 Hz cannot not be heard, such as via telephone. Formants are modified according to the spectral shape of the phonograph FRFs. An example is the disappearance of some formants and the development of new formants in figure 3.

Time domain: Singers reported an "echo" feedback effect during the recording process, which is also heard in the microphone recordings. In a way, the singer can foresee (better "forehear") how the voice will be shaped on the record, like by the reflected sound in a reverberant hall. To some extent, this helps the singer to control the sound and to adapt the voice to the medium. For today's singers, this effect causes some kind of "acoustic aura" of recordings using historic devices.

Another prominent change in timbre – apart from the increase of noise and the bandwidth limitation – is the strongly variable intensity of single notes and the sudden appearance of distortions in some notes whereas others do not exhibit such strong nonlinearities.

A second aspect of our research is to find out whether the observed voice timbre changes can be explained by the transfer characteristics of the historic devices and to what extent singers have modified or adopted their singing style to the recording situation.

Current research

Combined analysis and synthesis

To investigate the impact of transfer functions on the voice timbre and performance at early recording times a mixed approach of voice and device analysis and joint synthesis of the complete recording and reproduction chain is currently implemented. Whereas recording and reproduction device properties are not completely known today, part of these signal paths can be modeled based upon measured and calculated transfer functions [10].

Simulation of resonance influence

A "Max4Live"(M4L)-Patch (based on the conjunction of the Digital Audio Workstation (DAW) Ableton Live with the MAX/MSP audio processing system) has been developed in order to simulate and investigate the interaction of vocal formants and gramophone resonances under occurrence of vibrato. A signal flow chart of this set-up is shown in figure 4.

For this purpose, artificial vowels with a controllable amount and rate of vibrato and a user-definable fundamental frequency are created by the patch. The spectral characteristics of both the recording- and playback-chain of our measured gramophones [10] are then imposed on these vowels by convolution with the respective impulse responses.

The artificial vowels are created by feeding the signal of a tunable sawtooth oscillator through cascaded resonant formant-filters. These are modeled as standard biquad filters in peak-notch-mode. The formant frequencies are user-switchable for the choice of a desired vowel, currently a, e, i, o, u and ə (schwa). Two formants per vowel are used and set to values from literature [8]. A broad bandpass filter is added for increased realism. Vibrato is achieved by modulating the fundamental frequency of the sawtooth oscillator (carrier) with a sine oscillator (modulator). Hence, the frequency of the modulator determines

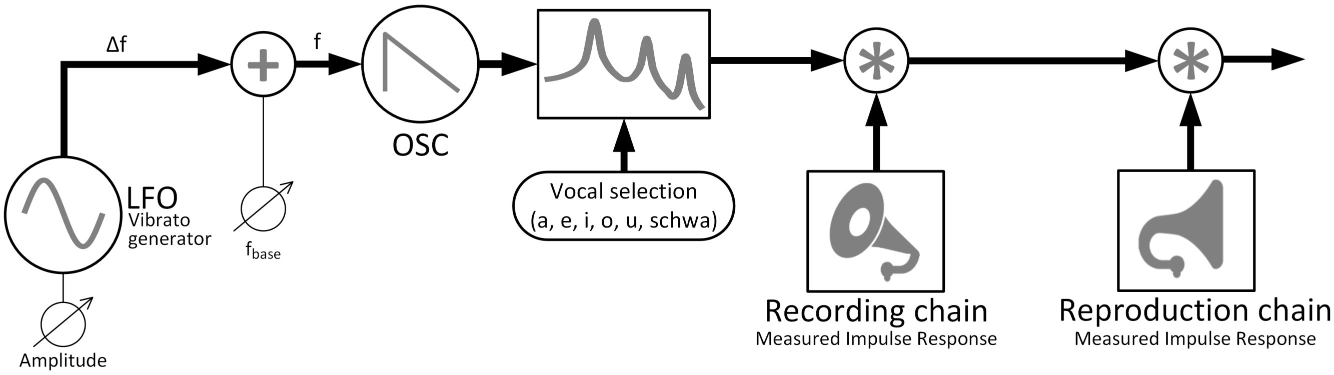

Figure 4: Signal flow of simulation set-up.

the vibrato speed while its amplitude yields the amount or range of the vibrato in Hz. In order to make this range user-selectable in "cent" rather than Hertz, which ensures a consistent vibrato within all registers, the amplitude of the modulator is derived from the base frequency of the carrier. The complex vibrato of the human voice with fluctuating pitch, level and timbre is here reduced to the first parameter, although level fluctuations still arise due to the influence of the formant filters. This reduction of complexity is feasible since we are interested in the changes of level (and timbre) which occur from elements later in the chain.

The impulse responses (IRs) taken from measurements of real gramophone recording- and playback setups (in the work of Tobias Weege [10]) are convolved with these artificial voice signals using Ableton Live's integrated convolution plug-in "Convolution Reverb". The results are equivalent to offline-convolutions done e.g. in MATLAB. By combining two instances of Convolution Reverb, it is possible to either investigate the recording section, the playback section or both in combination.

The spectrum of the signal before and after the convolution is compared visually in an overlay of two channels on the software frequency analyzer Voxengo "SPAN" (a VST-plug-in). With the vibrato set to a slow rate and with the analyzer set to fast time constants, the time-varying gain and attenuation of harmonics originating from the vibrato in the voice signal interacting with the resonances of the gramophone can be observed. SPAN also displays the current overall peak and RMS levels for both channels. Here it can be shown directly that the frequency modulation of the input results in an amplitude modulation of the output analogous to FM-Radio-transmission, when the fundamental or the harmonics of the vocal signal wander through sharp transitions of the gramophone frequency response for certain settings of the fundamental.

Influence on vibrato

The same effect can be observed when actual modern and undistorted voice recordings made by our group are convolved with the gramophone model IRs used here. Several short voice samples have been chosen to demonstrate this effect. Depending on the singer's register (soprano, tenor or bass) and singing style (with/without singer's formant), vowels are affected differently by the convolution. Apparently, this is due to the variable combinations of harmonics, vowel formants and gramophone resonances.

In order to approach the effect systematically, we selected voice examples of distinct vowels at a constant pitch that where audibly affected by a strong amplitude modulation compared to the original. We found that all these examples always showed a significant presence of vibrato. It is observed that the amplitude modulation can be twice as fast as the original vibrato for vowels of certain pitches, resulting in a doubling of the perceived vibrato rate. Since the variation of pitch by vibrato varies all harmonics of the regarded voice signal, all of them are modified differently by the device FRFs, thus they interact in a complex way. A combined "vowel signal response" gives the output level for a specific vowel with all its harmonics as a function of the fundamental pitch. This response can be strongly different from the device FRF and exhibits narrow minima and maxima as shown in figure 5. With a base pitch residing at such a local

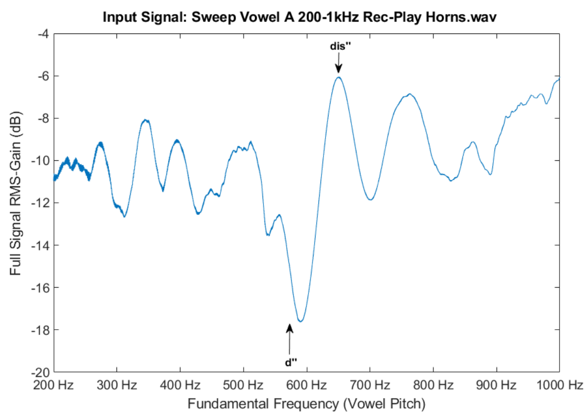

Figure 5: Impact of device FRFs on level of swept artificial vowel /a/ without vibrato

extremum, vibrato causes a significant change in overall level towards both sides of the center frequency, which doubles the (amplitude-) vibrato modulation rate.

In another example we used a sweep with vibrato sung by a tenor, using a different vowel each time. The amount of vibrato was kept approximately constant over the sweep. The effect of the FRFs is shown in figure 6. The amplitude of the convolved voice signal shows almost no vibrato at lower pitch and at high pitch but strong and even amplified signal for some pitches in the mid-rang of the sweep. This clearly shows the dependence of the effect on the fundamental frequency of vowels, where a certain range of pitches undergoes modulation while others remain unaffected.

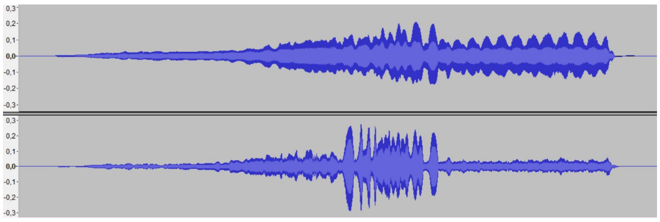

Figure 6: Impact of device FRFs on swept vowel /a/ with vibrato: microphone signal without (top) and with convolved device FRF (bottom)

Future work

It was observed that the recording and reproduction FRFs and artefacts strongly depend on the dynamics of the input signal and on the mechanical properties of the device. The latter are the jitter induced by the variable angular velocity of the mechanism, wow and flutter due to the the non-perfect shape of the cylinder which can be more oval than round, and of course the skills of the operator who handles the devices. These parameters have been estimated and modeled earlier [9] and shall be investigated more closely with respect to their impact on the timbre of voice recordings.

The modification of the singer's formants by historic recording devices have indicated the need for a closer look at the sensitivity of voice timbre to the transfer functions of the recording and reproduction devices. A perceptual study could help to judge whether the timbre changes would be rated as negligible, detrimental or advantageous for the judgment of singers' voices.

Regarding the usage of vibrato, it shall be further investigated to which extent a singers vibrato could have been technically unsuitable for recording due to excessive amplitude modulation as an artifact. Deeper analysis of test signals from vocal models and actual vocal recordings should be accomplished for investigating this.

Another future step will be the reduction of the devices' influences using inverse filtering to estimate the historic singing voice qualities and performance practice.

Acknowledgements

This work has been supported by DFG grant KO 2874/7-1 for the project "Technologies of Singing (TdS)". The authors thank George Brock Nannestad for his valuable discussions on part of the investigations.

References

[1] Habasinska, D., Weege, T. A., Kob, M.: The influence of early recording and playing devices on the singing practice, as a research question of the project "Technologies of Singing". Presentation at 18. Diskografentag, May 12-13 2017, Hannover

[2] Holtsträter, K.: "Der Crooner, das unbekannte Wesen". In: Sabine Meine, Nina Noeske (Hrsg.): „Musik und Popularität. Aspekte zu einer Kulturgeschichte zwischen 1500 und heute". Reihe „Populäre Kultur und Musik", Waxmann 2011, Band 2

[3] Kob, M.: Stimmaufnahmen auf Tonträgern. Presentation at 17. Vortrags- und Gesprächsrunde "Stimmerkrankungen bei Schauspielern und Sängern", November 18-18 2016, Wien

[4] Martensen K., Zakharchuk P., Kob M., Grotjahn R.: Phonograph und Gesangsstimme – Untersuchungen zur Akustik früher Aufzeichnungs- und Abspielgeräte. Proceedings DAGA 2015, March 16-19 2015, Nürnberg, 1429-1432

[5] Martensen, K.: The construction of a media voice: Findings in the T.A. Edison Historical Site, New Jersey. Historical Performance, 2018, in press

[6] Pirch, P.: On the Use and Acoustical Characteristics of Mechanically Augmented Musical Instruments and Early Recording Devices. Master's thesis, 20.9.2015, HfM Detmold

[7] Proctor, M., Bresch, E., Byrd, D.; Nayak K., Narayanan, S. (2013): "Paralinguistic mechanisms of production in human 'beatboxing': A real-time magnetic resonance imaging study", Journal of the Acoustical Society of America 133(2), 1043-1054

[8] Sendlmeier, W. F., Seebode, J.: Formantkarten des deutschen Vokalsystems. URL: `http://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf`, aufgerufen 10.3.2018

[9] Välimäki, V., González, S., Kimmelma, O., Parviainen, J.: Digital Audio Antiquing-Signal Processing Methods for Imitating the Sound Quality of Historical Recordings. JAES 56 (3) 2008, 115-139

[10] Weege, T., Habasinska, D., Kob, M.: Influence of Early Recording and Playing Devices on Musical Sound: FRF Measurements of Horn, Soundbox and Tonearm. Proceedings DAGA 2018, March 19-22 2018, Munich, in print