euronoise 2021
25th to 27th of October
Madeira, Portugal - online

Time-variant signal manipulation using frame multipliers

Christian H. Kasess[1], Thomas Maly[2], Peter Balazs[1], Wolfgang Kreuzer[1]

[1] Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria
{Christian.Kasess@oeaw.ac.at, Peter.Balazs@oeaw.ac.at, Wolfgang.Kreuzer@oeaw.ac.at}
[2] Institute of Transportation, TU Wien, Vienna, Austria
{thomas.maly@tuwien.ac.at}

Abstract
To better understand the relation between annoyance and acoustic properties of environmental noise, laboratory studies with defined noise samples are often performed. To reduce measurement effort, sample generation through manipulation is an important approach. In some cases, it is necessary to use time-variant filters, e.g., to extract signal components whose frequency content changes over time. The method presented here is based on frame multipliers. This allows the definition of such filters directly in the time-frequency plane. The versatility of this approach is illustrated using curve squeal of railway vehicles, which can comprise narrow-band components affected by the Doppler shift but also broadband noise which may vary in intensity and frequency range. Multipliers are used to extract the relevant signal portions and to combine defined pass-by noise samples with the extracted squeal noises to produce well defined test conditions.

Keywords: time-varying filtering, frames, multipliers, curve squeal, test stimuli.

1 Introduction

Due to the increased burden caused by environmental noise, research on the perception of such noise and on mitigation measures is an important topic. Many laboratory studies have been performed focusing on the relation between acoustic parameters, psychoacoustic parameters, and annoyance ratings under various conditions. Examples for such studies are the effect of noise barriers for road traffic [1], railway traffic [2], or the effect of different source conditions such as varying wheel roughness [3], the effect of rumble strips [4], or different variants of high speed trains [5], among many others.

In the majority of cases, time consuming and costly measurements are the basis of such investigations. Sometimes, if test data is difficult to acquire or not available from measurements, sound samples are modified using simple time-independent filtering or scaling. For instance, in a previous work the spectral effect of a noise barrier was considered based on the results of numerical simulations [2]. As the sound sample consisted of stationary portions of a train pass-by, the effect of the noise barrier was assumed to be constant during the duration of the sound sample although the barrier's effect on the contribution of each wheel is time-dependent due to the motion. Clearly, this is a simplification and approaches such as time windowing or spectral filtering may often not be sufficient, e.g. when clearly audible non-stationary components are present.

One example requiring more sophisticated approaches are rumble strips, where test data of different strip parameters was generated using a combination of filtering and partially synthesized audio data [4]. If, however, there is no sufficiently good model to generate artificial sound samples time-dependent signal processing may be an option for the specific purpose.

One such method are Gabor multipliers [6] which act directly in the time-frequency-domain and are a particular way to implement time-varying filters [7]. To allow other kinds of time-frequency transforms one reaches the more general frame multipliers [8]. The concept of such systems for signal modification is very simple, analysis (i.e. transformation into the time-frequency-domain) is followed by point-wise multiplication with a filter mask (i.e. the multiplier) and re-synthesis. While those operators are interesting from a mathematical point of view [9][10][11], they also have applications in acoustics and signal processing [12][13][14][15][16]. A graphical user interface to apply such multipliers was implemented in the Large Time-Frequency Analysis Toolbox (LTFAT [17], http://ltfat.org/) by the name of MULACLAB [18][19] and available for Octave and MATLAB.

Here, railway curve squeal is used as an example to illustrate the usefulness of time-dependent filtering via frame multipliers. Curve squeal is a railway specific effect that occurs in tight curves. There are two different mechanisms producing curve squeal. In curves the wheels may periodically roll and then slide lateral leading to a jerking motion (stick-slip effect [20]). This leads to an excitation of wheel modes which in turn results in a narrow-band sound emission. If the propagation from the wheel to the microphone were time-independent, a simple filtering approach may suffice to isolate the squeal. However, the wheel is moving with respect to the observer and thus a time-dependent frequency shift known as the Doppler-effect occurs, resulting in a more complex time-frequency pattern. The second mechanism causing curve squeal is when the wheel flange gets in contact with the rail. The so generated noise is also called flanging noise or flanging squeal and is typically a broad-band variant of curve squeal. Both variants will be discussed in this work.

The structure of the manuscript is as follows. First, frames and frame multipliers will be explained in some detail. Then, an overview of MULACLAB (the implementation in the LTFAT [18]) will be provided. Some details on the measurements are given and then different cases of combining curve squeal with regular pass-bys are demonstrated and discussed. The basic principles and the feasibility were presented in [21] using tonal squeal. Here, this is extended to broad-band squeal. Since this previous publication is in German and to keep the manuscript self-contained the main principles are also presented here.

2 Methods

2.1 Frames

One of the most important tasks in signal processing is the representation of a continuous signal by discrete coefficients. A typical example is the Shannon sampling theorem [22], where an analog signal is represented by discrete sampling values. In order to achieve this, one searches for atoms ψ_k such that

$$f(t) = \sum_{k=1}^{\infty} c_k \psi_k(t) \tag{1}$$

for any reasonable signal f, e.g. $f \in L^2(\mathbb{R})$, i.e. having finite energy.

The classical approach is to use orthonormal bases for ψ_k, but redundant representations have found a lot of applications recently [23]. To still guarantee perfect reconstruction one introduces the following concept:

A sequence of vectors $\Psi = (\psi_k)_{k \in K}$ in a Hilbert space H is called a *frame*, if constants $A, B > 0$ exist, such that

$$A \cdot \|f\|^2 \leq \sum_{k \in K} \left| \langle f, \psi_k \rangle \right|^2 \leq B \cdot \|f\|^2 . \tag{2}$$

Frames allow non-unique, redundant representations, but still allow perfect reconstruction. This is achieved by using a dual frame, which always exist.

In this manuscript we use a discrete Gabor transform (DGT), i.e., a frame consisting of time- and frequency shifted copies of a window g:

$$\psi_{\mathrm{m,n}}(t) = g(t-am)\mathrm{e}^{2\pi bn}. \quad (2)$$

Thus, this is a short-term Fourier transform where *m* and *n* denote the temporal and spectral shift in the time-frequency plane, respectively. With $ab < 1$ and e.g. $g(t)$ a Gaussian window function this transform fulfills the properties of a frame.

2.2 Multipliers

Multipliers are a way to define a manipulation directly in the transform domain. The analysis coefficients are multiplied by a fixed symbol and the resynthesized. So, for the symbol or mask $m = (\mathrm{m}_k)_{k\in K}$ and sequences $\Psi = (\psi_k)_{k\in K'}$, $\Phi = (\phi_k)_{k\in K}$ we define a frame multiplier as

$$\mathbf{M}_{m,\Psi,\Phi}\; f = \sum_{k\in K} m_k \langle f, \phi_k \rangle \psi_k \quad . \quad (3)$$

Here two different frames can be used. Most often a frame and its dual are used so that the multiplication with symbol $m_k \equiv 1$ corresponds to the identity.

In the MATLAB/Octave *Large Time-Frequency Analysis Toolbox (LTFAT)* frames are implemented in an object-oriented approach, where any transform can be put in this context. We use a Gabor transform implemented by `dgtreal` in LTFAT. The application of frame multipliers use `framemul`, but, in particular, there is a special graphical user interface for those operators.

2.3 MULACLAB

MULACLAB is a graphical user interface for using multipliers as described above. It allows to import wav-files for which the frame analysis coefficients representation is calculated and displayed. In our case this is a time-frequency representation based on a Gabor transform, although essentially all important time-frequency representations implemented in the LTFAT can be used such as Gabor-frames [24], Erblets [25], discrete wavelet transform [26] and others and the parameters can be defined in a wide range.

On the basis of the representation, e.g. the spectrogram in the case of Gabor-frames, time-frequency regions which contain the signal of interest are defined. For the definition of regions different operations are possible. First, regions can be manually defined by drawing the appropriate outlines. Second, using the "wand" tool a region around a defined point in the TF-plane can be generated using simple region growing. The dynamic range can be defined in dB. Third, a simple sub-band can be defined which is similar to a time-invariant filter.

Multiple regions can be defined using either of the above-mentioned methods which are combined with the already existing region(s) using different set operations: union, intersection, and difference. Furthermore, multiple layers can be defined which are then applied sequentially. By default, the marked region in a layer is cut out thus producing a hole in the time-frequency plane. Each multiplier defined in a layer can also be "inverted" by using the symbol $(1-m_k)$ in order to remove the background around the region and thus isolate the desired signal component. The selection as well as the TF-representation can be stored and imported directly.

After applying the multiplier, the modified signal can be played back and exported to a wav-file.

2.4 Data acquisition and processing

The usefulness of binary multipliers, i.e., $m_k \in \{0,1\}$ for our purposes will be illustrated using examples of curve squeal. The data were acquired next to a railway curve in 25 m and 50 m distance from the track using microphones. At 25 m a head-and-torso simulator was also placed which was used for binaural recordings, however, these data are not shown here. From the microphone measurement at 25 m two sound samples of length 5.8 s were extracted to demonstrate the approach. First, a suitable single instance of a narrow-band squeal was identified in a pass-by of a cargo train with a speed of 80 km/h. To illustrate the broad-band variant, an occurrence for a passenger train was chosen. Different clean pass-by recordings (i.e. train pass-bys without any apparent tonal squeal or flanging noise) were used as a basis.
For the remainder of the paper the samples will be referred to as tonal or broad-band squeal recording and clean recording, respectively.
The Gabor frame for the time-frequency representation used a Hann window and a window length of 1024 samples with a hop size of 256 samples and 2048 frequency bins.

3 Results

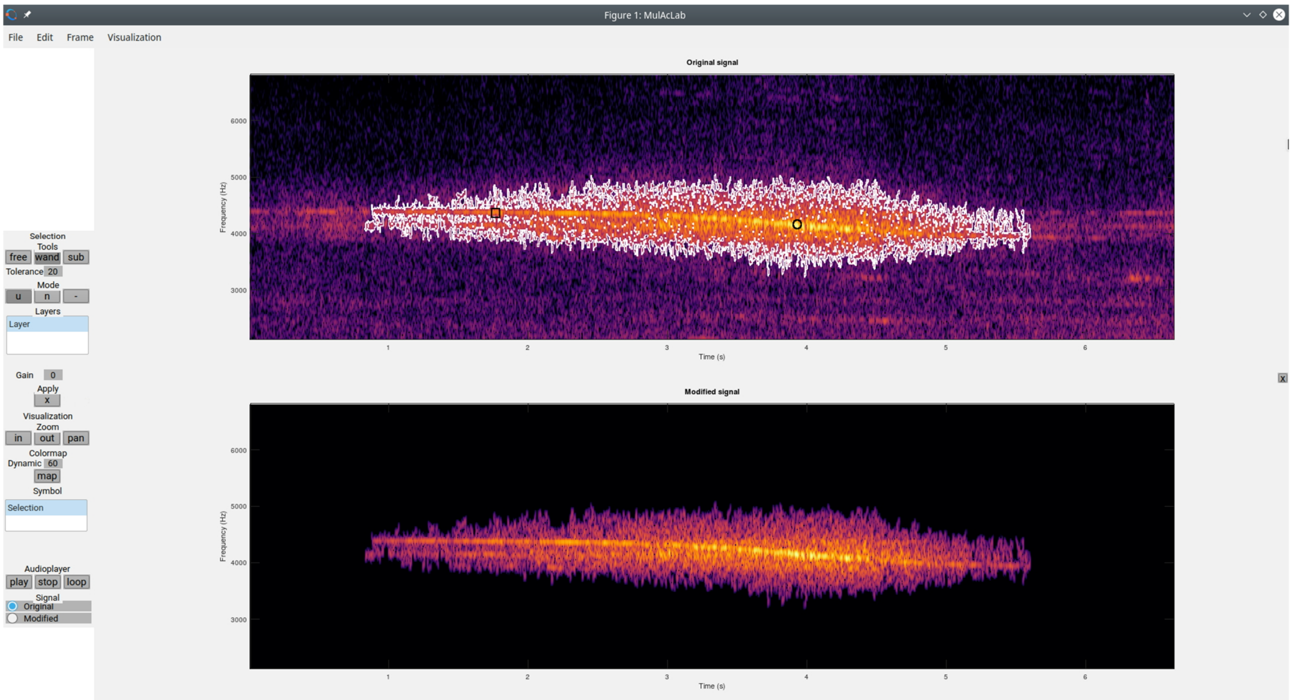

Figure 1 – Screenshot of MULACLAB in Octave. The upper panel shows the time-frequency representation of the original signal and the selected region. The black circle and rectangle indicate the first and second seed region for the region-growing algorithm. The lower panel shows the signal after applying the multiplier.

Figure 1 shows the DGT of the original signal where the narrow-band curve squeal around 4 kHz is clearly visible. The Doppler-shift can be nicely seen and thus it is clear that simple bandpass filtering will not suffice to isolate the squealing noise.

The crucial step in this setting is the definition of a suitable cutoff-region which is difficult due to the noisy background. Two main approaches could be used: manually drawn regions or a region growing approach. Manually drawn regions will not be considered here but were illustrated in [21].
Using the "wand" tool of MULACLAB a region around the component of interest is generated (Figure 1). Clearly, the starting point and the dynamic range are essential to define the size of the region. Here, a range of 25 dB was chosen and a starting point at the maximum of the squeal noise (circle in upper panel of Figure 1). In a second step, another starting point was chosen with 20 dB to capture the initial phase of the squeal.
Either approach requires some experience and intuition about the signal in question. The sub-band definition which approximates a time-invariant filter was not suitable for the signals used here.

When using the "inverse multiplier" the signal component of interest can be extracted (Figure 1, lower panel). Due to the perfect reconstruction property the sum of the signal constructed with the multiplier and the inverse multiplier is, up to a numerical error, equal to the original signal. This is a big advantage of the linear multiplier approach.

When listening to either of the modified signals, i.e. the squeal alone or the squeal cut out, no musical noise or distortion was audible. This may, however, also be a consequence of the non-stationary nature of the squeal and train noise which may make it hard to identify any phase distortions in the signal. For the manual mask a slight remnant of the squeal could be perceived.

To generate pass-by noise samples with curve squeal from recordings without squeal, the idea is to add the squeal signal to the pass-by sample. There are two ways this can be done. The signal can be added with or without cutting out the time-frequency tile in the clean signal where the squeal noise will be added with the former approach being the more adequate as it avoids the summation of sound energies of the underlying pass-by noise of two recordings in that region. If the squeal is sufficiently high in level, this effect will probably be negligible. However, e.g. when the magnitude is modified cutting out the signal portion before combining the samples may become important.

In [21] the combination was done using the original squeal mask to cut out the pass-by noise before combining the sounds. Here, a slightly different approach is introduced which is more flexible. Instead of using this pre-defined mask, the two time-frequency representations of the squeal and the clean signal are compared and a binary mask is generated with regions of one where the squeal sample is higher in amplitude and of zero where the clean sample is larger. To avoid too noisy a mask, i.e. small speckles of ones and zeros, a further step is included. Applying morphological closing and opening operations where holes (small regions of zeros surrounded by only ones) are closed and small regions of ones surrounded by zeros are set to zero. The mask is applied to the squeal and the inverse mask is applied to the clean signal and these two are combined. A further advantage of this approach is that modifications e.g. of the frequency are implicitly taken into account which would be difficult using a predefined mask.

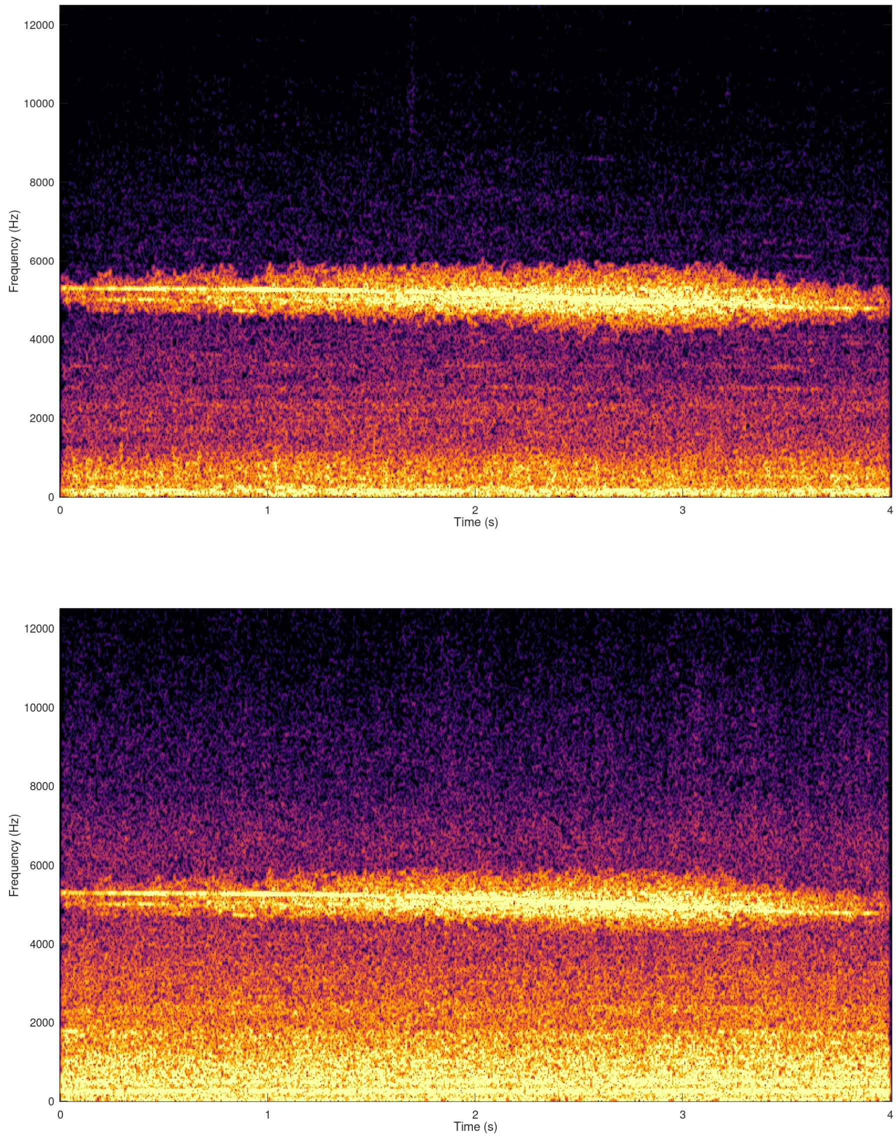

Figure 2 – Tonal squeal combined with the rolling noise of a passenger train (upper panel) and a cargo train with higher rolling noise (lower panel). The squeal has been frequency shifted to a center frequency of 5 kHz before its application using a phase vocoder.

Figure 2 shows squeal-free pass-bys of a passenger train (upper panel) and a cargo train (lower panel) where the squeal was added using the procedure described above. The cargo train had a much higher rolling noise thus the squeal is not as prominent in the time-frequency plane. Note, that the tonal squeal in this figure was further manipulated by changing the center frequency to 5 kHz using a phase vocoder [27].

Clearly, in this setting there is no objective way to evaluate the signals for the two methods as there is no clean reference signal to compare to. A main issue for the comparison is that the background noise of a pass-by will always vary, so two samples will never be identical. In [21] a rating was performed by an expert listener (the second author) on how realistic the signals are and whether distortions are audible. The listener who is experienced in the analysis of curve squeal was asked to identify which samples he thought were modified. Different samples were prepared by the first author comprising the original pass-by with squeal and two combinations of a section without squeal from the same pass-by and of the extracted squeal (with and without a-priori cutting out of the clean signal portion). Furthermore, two masks were used: one based on region growing and a manually drawn. The expert listener was not aware which sample was the original squeal and which one was modified.

Briefly, as expected, due to additional noise components the listener was able to distinguish the two different recordings underlying the test, but he was not able to point out which samples were modified. All samples were judged to be not modified. However, the added squeal was perceived as slightly more dominant than the squeal in the unmodified recordings which may be a consequence of different background noise in the recordings.

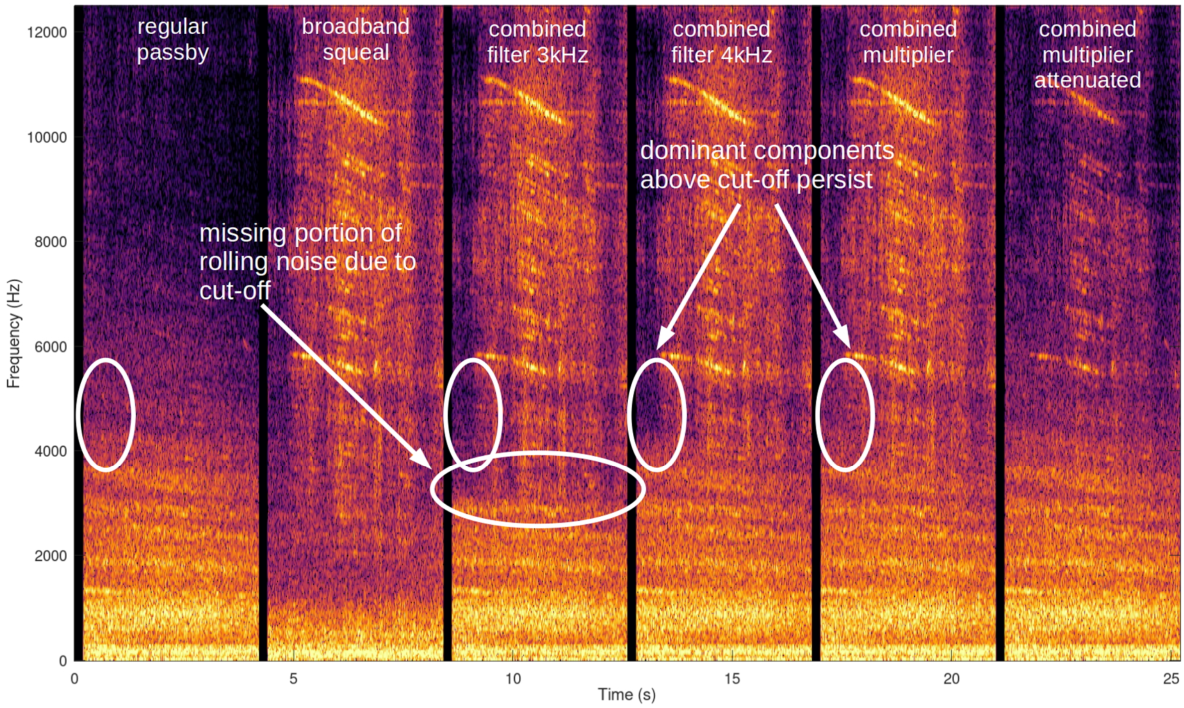

Figure 3 – Different approaches of combining broad-band squeal (flanging noise, second signal) with a clean rolling noise sample (first signal). The third and fourth signal uses a simple filterbank of a high-pass and low-pass filter with a cut-off frequency of 3 kHz and 4 kHz, respectively. The fifth signal shows the results for the multiplier approach. For the sixth signal, the squeal was attenuated before combining it with the clean signal.

Applying this method to broad-band stimuli requires some modification. Figure 3 shows first a clean sample and second a broadband sample. Naturally, it is not possible to clearly identify the regions of interest. Thus, no manual mask is generated for this application. However, it is necessary to define a transition zone below which only the rolling noise of the clean reference sample will be present. In this example, the flanging noise occurred during the pass-by of an otherwise relatively silent train. The clean sample is of a louder cargo train. Here, a cut-off of 1.5 kHz was applied to the time-frequency representation of the squeal sample. Then, a binary mask was produced using the procedure described above by comparing the clean and the squeal sample. After the application of the mask the two signals were combined to result in the fifth signal in Figure 3. For comparison, the two samples were also combined using conventional filters with a cut-off at 3 and 4 kHz. For 3 kHz it is clear that major portions of the clean rolling noise are gone which results in a muffled sound at the beginning of the sample where the squeal is very weak. For 4 kHz the situation looks much better. However, the results are clearly highly dependent on the choice of cut-off whereas for the multiplier one only needs to take care that no low-frequency portion of the rolling noise of the squeal sample dominates the clean sample.

Thus, it is possible to generate a highly controlled set of stimuli for perception tests. Furthermore, when modifying the squeal by, e.g., attenuating the signal or changing the frequency content by using a frequency dependent multiplier for example, the multiplier approach still works whereas for the filter a different cut-off would need to be defined (sixth sample in Figure 3).

4 Conclusions

Summarizing, the presented examples illustrate the versatility of the multiplier approach using tonal and broad-band curve squeal. As also previously reported, the generated samples sounded natural and no musical noise or other distortions were audible. The tool MULACLAB which is part of the LTFAT was shown to provide a flexible tool to combine masks generated using either a purely manual approach or region growing.

The method was illustrated to enable the generation of sound samples which would be difficult to acquire using only measurements. The main objective was to produce samples which were used in a perceptual experiment evaluating the effects of different variants of curve squeal on annoyance. The method allows to produce a set of systematic variations of specific squeal noises.
For this, a range of modifications of the squeal before combining it with clean pass-bys was necessary. Here, results of pitch shifting a tonal squeal as well as simple attenuation were illustrated, however other modifications such as changing the spectral tilt of broad-band squeal can be easily implemented.
Comparing the spectro-temporal amplitudes of the squeal samples and the clean signal in order to generate a mask for combining the two signals provides a method that does not require a large amount of manual adjustments. However, as illustrated in the examples presented some knowledge of the signal properties is required to set up the basic parameters.

Further processing can also be performed generating e.g. a spatialized tonal squeal sample based on head-related transfer function. This sample can be included into a binaural recording using a head-and-torso simulator. For this, the actual position of the squealing wheel over time needs to be considered. Initial judgments of the expert show that adding the squeal sounds natural including the spatial impression. However, probably due to the delay and amplitude changes between the channels, the opposite operation, i.e. removing an unwanted signal may lead to some audible effects in the binaural case. This is a matter of further investigation.

Acknowledgements

This work was in part supported by the Austrian Research Promotion Agency (FFG, project 860523), the Federal Ministry for Climate Action, Environment, Energy, Mobility Innovation and Technology, and the Austrian Federal Railways (ÖBB) as well as the project P 34624 of the Austrian Science Fund (FWF).

References

[1] Nilsson, M. E.; Andéhn M.; Leśna, P. Evaluating road-side noise barriers using an annoyance-reduction criterion. *The Journal of the Acoustical Society of America*, 124(6), 2008, 3561–3567.

[2] Kasess, C. H.; Maly, T.; Majdak P.; Waubke, H. The relation between psychoacoustical factors and annoyance under different noise reduction conditions for railway noise. *The Journal of the Acoustical Society of America*, 141(5), 2017, 3151-3163, 2017.

[3] Kasess, C. H.; Noll, A.; Majdak, P.; Waubke, H. Effect of train type on annoyance and acoustic features of the rolling noise. *The Journal of the Acoustical Society of America*, 134(2), 2013, 1071–1081.

[4] Kasess, C. H.; Maly, T.; Kluger-Eigl, W.; Waubke, H. Psychoacoustic evaluation of different rumble strip designs. *Proceedings of the Internoise 2019*, Madrid, 2019, In CD–ROM

[5] Vos, J. Annoyance caused by the sounds of a magnetic levitation train. *The Journal of the Acoustical Society of America*, 115(4), 2004, 1597–1608.

[6] Feichtinger, H. G.; Nowak, K. A first survey of Gabor multipliers, *Advances in Gabor Analysis. Applied and Numerical Harmonic Analysis*, Birkhäuser, Boston, MA, 2003.

[7] Matz, G.; Hlawatsch, F. Linear Time-Frequency Filters: On-line Algorithms and Applications. *Application in Time-Frequency Signal Processing*, B. Raton (FL): CRC Press, 2002.

[8] Balazs, P. Basic definition and properties of Bessel multipliers. *J. Math. Anal. Appl.*, 325(1), 2007, 571–585.

[9] Stoeva, D. T.; Balazs, P. Invertibility of multipliers. *Appl. Comput. Harmon. Anal.*, 33(2), 2012, 292–299.

[10] Gröchenig, K. Representation and approximation of pseudodifferential operators by sums of Gabor multipliers. *Appl. Anal.*, 90(3-4), 2010, 385–401.

[11] Balazs, P.; Stoeva, D. Representation of the inverse of a multiplier. *J. Math. Anal. Appl.*, 422, 2015, 981–994.

[12] Brown G. J.; Cooke, M. Computational auditory scene analysis. *Computer Speech & Language*, 8(4), 1994, 297–336.

[13] Majdak, P.; Balazs, P.; Kreuzer, W.; Dörfler, M. A time-frequency method for increasing the signal-to-noise ratio in system identification with exponential sweeps. *Proceedings of the 36th International Conference on Acoustics, Speech and Signal Processing, ICASSP 2011*, Prag, 2011, In CD-ROM.

[14] Balazs, P.; Laback, B.; Eckel, G.; Deutsch, W. A. Time-frequency sparsity by removing perceptually irrelevant components using a simple model of simultaneous masking, *IEEE Transactions on Audio, Speech and Language Processing*, 18(1), 2010, 34–49.

[15] Olivero, A.; Torrésani, B.; Kronland-Martinet, R. A class of algorithms for time-frequency multiplier estimation. *IEEE T. Audio. Speech.*, 21(8), 2013, 1550–1559.

[16] Tauböck, G.; Rajbamshi, S.; Balazs, P.; Abreu, L. D. Random Gabor multipliers and compressive sensing, *Sampling Theory and Applications (SampTA) 2019*, Tallinn, 2019, In CD-ROM.

[17] Søndergaard, P.; Torrésani, B.; Balazs, P. The linear time frequency analysis toolbox, *Int. J. Wavelets Multi.*, 10(4), 2012.

[18] Průša, Z.; Søndergaard, P. L.; Holighaus, N.; Wiesmeyr, C.; Balazs, P. The Large Time-Frequency Analysis Toolbox 2.0. Sound, Music, and Motion, M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, Eds., Lecture Notes in Computer Science, pp. 419–442. Springer International Publishing, 2014.

[19] Stoeva, D.; Balazs, P. A survey on the unconditional convergence and the invertibility of multipliers with implementation. 2020.

[20] Thompson, D. Curve Squeal Noise. *Railway Noise and Vibration*, Elsevier, Oxford, 2009.

[21] Balazs, P.; Kasess, C.; Kreuzer, W.; Maly, T.; Průša, Z.; Jaillet, F. Anwendung von Rahmen-Multiplikatoren für die Extraktion von Kurvenquietschen von Zugsaufnahmen, e & i Elektrotechnik und Informationstechnik, 138, 2021, 206–211.

[22] Zayed, A. I. *Advances in Shannon's Sampling Theory*, CRC Press, 1993.

[23] Balazs, P.; Holighaus, N.; Necciari, T.; Stoeva, D. Frame theory for signal processing in psychoacoustics. *Excursions in Harmonic Analysis* Vol. 5, R. Balan, J. J. Benedetto, W. Czaja, and K. Okoudjou, Springer, 2017.

[24] Feichtinger, H. G.; Strohmer, T. *Gabor Analysis and Algorithms - Theory and Applications*, Birkhäuser Boston, 1998.

[25] Necciari, T.; Holighaus, N.; Balazs, P.; Průša, Z.; Majdak, P.; Derrien, O. Audlet filter banks: A versatile analysis/synthesis framework using auditory frequency scales. *Applied Sciences*, 8(1), 2018.

[26] Průša, Z.; Søndergaard, P. L.; Rajmic, P. Discrete Wavelet Transforms in the Large Time-Frequency Analysis Toolbox for Matlab/GNU Octave. *ACM Trans. Math. Softw.*, 42(4), 2016, 32:1–32:23.

[27] Průša, Z.; Holighaus, N. Phase vocoder done right. *Proceedings of 25th European signal processing conference (EUSIPCO-2017)*, Kos, 2017, 1006–1010.