Shiftr

Spectral pitch shifter for Voltage Modular



Shiftr changes the pitch of a signal by up to two octaves either way without changing its duration. It works on the signal's spectrum: it finds each partial, moves it to its new frequency, and keeps its phase continuous from one analysis frame to the next. It can keep the spectral envelope, and so the formants, where they were while the pitch moves, or move the envelope independently of the pitch. It is stereo, with feedback, a spectral freeze and a dry path delayed to line up with the shifted signal.

This manual describes every control and jack, and gives measurements taken from the module's own DSP code.

Contents

Overview 2
1 · SHIFT 4
2 · FORMANT 6
3 · FEEDBACK and FREEZE 8
4 · I/O 10
5 · The display 10
6 · How the shift works 12
7 · Control ranges 13

Overview

> **shift** /ʃɪft/
>
> **1** *verb* move, or cause to move, from one place or position to another.
> **2** *noun* a change of frequency applied equally, in ratio, to every component of a signal.

Shiftr multiplies the frequency of every partial in its input by the same ratio, set in semitones. A note shifted up 7 semitones comes out a fifth higher, at the same speed. The input's spectral envelope can be kept in place, so that a voice shifted up keeps the resonances of the throat and mouth that make it sound like the same voice, or moved on its own, so that the voice changes character without changing note.

What it is

Shiftr is two identical channels, left and right, each analysing its input in overlapping frames, shifting the frame's spectrum, and resynthesising it. Several of its controls start where they have no effect:

Control	What it does	No effect at
PITCH, **FINE**	the shift, in semitones and cents	0
AMT (PITCH)	scales the PITCH MOD input	its centre, or 0 V
PRESERVE	holds the spectral envelope in place	off
FORMANT	moves the spectral envelope	its centre, 0
AMT (FORMANT, FEEDBACK, MIX)	scales each CV input	its centre, or 0 V
FEEDBACK	returns the output to the input	fully left
FREEZE, **GATE**	holds the current spectrum	off, or below 0.5 V
MODE	how FREEZE holds it	while not frozen
SPREAD	detunes the two channels apart	fully left

With every control where it starts, Shiftr passes its input to its outputs unchanged in pitch and level, delayed by 53.3 ms.

Signal path

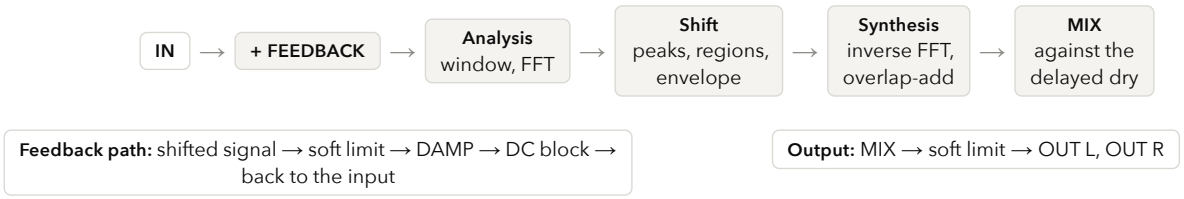


Figure 1. One channel. The other is the same. IN R is IN L when nothing is patched into it.

Latency

The shifted signal leaves Shiftr one window and a quarter after it arrives. The dry signal is delayed by the same amount, so that MIX blends the same moment of both. The delay depends on **WINDOW**:

WINDOW	Delay	Frame interval	Frequency resolution
1K, 1024 samples	1280 samples, 26.7 ms	256 samples, 5.3 ms	46.9 Hz
2K, 2048 samples	2560 samples, 53.3 ms	512 samples, 10.7 ms	23.4 Hz
4K, 4096 samples	5120 samples, 106.7 ms	1024 samples, 21.3 ms	11.7 Hz

The frequency resolution is the spacing of the analysis's frequency bins at 48 kHz. Two partials closer together than about twice that spacing are not separated.

The face



Figure 2. The face, by section. Each number is the section of this manual that describes it.

1 **SHIFT**	the pitch, the window, attacks	2 **FORMANT**	the spectral envelope
3 **FEEDBACK, FREEZE**	feedback and damping; freeze and its mode	4 **I/O**	inputs, MIX, SPREAD, outputs
5 **Display**	the current frame		

Each control with a CV input has a small **AMT** knob beside it and the jack below the AMT knob, joined to it by dotted lines. Every input reads 0 V when nothing is patched into it.

1 · SHIFT



(1) **PITCH**	−24 to +24 semitones	(2) **AMT**	PITCH MOD's amount
(3) **PITCH MOD**	a CV on the pitch	(4) **FINE**	−100 to +100 cents
(5) **WINDOW**	1K, 2K, 4K samples	(6) **TRANSIENT**	keeps attacks in place

PITCH sets the shift in whole semitones, from −24 to +24. It moves in steps of one semitone and starts at 0. Its tooltip reads `Pitch: +7 semitones`.

FINE adds −100 to +100 cents. It starts at its centre, 0.

PITCH MOD adds a control voltage to the shift, through its **AMT** knob. AMT has no effect at its centre, where it starts. Turned right, a positive voltage raises the pitch; turned left, it lowers it. At either end the input is 1 V per octave, so a keyboard's or a sequencer's pitch CV patched here, with AMT fully right, plays Shiftr's output in tune. Between the centre and the ends the amount follows the square of AMT's travel (Figure 3), so the part of the knob near its centre gives the small amounts used for vibrato. AMT's tooltip gives the amount in cents or semitones per volt, as `Pitch mod: +18.0 cents a volt`.

AMT	Per volt	A ±5 V LFO swings the pitch
5%	3 cents	±15 cents
10%	12 cents	±60 cents
20%	48 cents	±2.4 semitones
50%	3 semitones	±15 semitones, held to ±24 in all
100%	12 semitones	the whole range

PITCH, FINE and PITCH MOD are added together, and the total is held between −24 and +24 semitones. The shift is taken up once per frame: every 10.7 ms at 2K. A fast LFO on PITCH MOD is therefore heard in steps of that length.

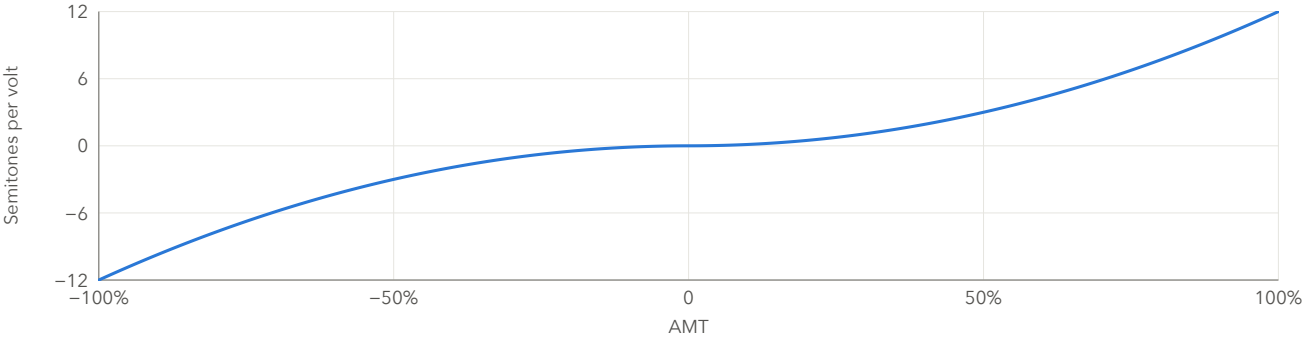


Figure 3. PITCH MOD's amount against AMT: 12 × AMT × |AMT| semitones per volt.

WINDOW sets the length of the analysis frame: 1K, 2K or 4K samples. It starts at 2K. A longer window separates partials that are closer together, which suits low notes, chords and pads, and spreads each attack over a longer time. A shorter window follows fast playing and percussion more closely, separates fewer partials, and has less delay (see Latency in the Overview). Its tooltip gives the window and the delay, as `Window: 2048 samples; latency 53.3 ms`.

A change of WINDOW fades the shifted signal out over 256 samples, waits while the new window fills with input, and fades it in again over 256 samples: 0.06 s in all from 2K to 1K, 0.20 s from 2K to 4K. The dry signal carries on through the change, and its delay moves to the new window's over 256 samples while the shifted signal is silent.

TRANSIENT keeps attacks in place. Shiftr carries each partial's phase on from frame to frame, which keeps a held note continuous. For a sudden attack those carried phases are wrong: each of the overlapping frames that hold the attack then rebuilds it at a different time, and it is heard spread over up to a window, much of it ahead of where it was played (Figure 4).

When the input's high-frequency content rises by TRANSIENT's threshold from one frame to the next, Shiftr marks an attack, and for the next four frames, those that hold it, it works out where in each frame the rising energy is. In each of those frames where the attack lies within one frame interval of the centre, one or two of them, each partial whose level has risen by the threshold since before the attack takes its phases from the input instead, adjusted for the shift so that the attack stays where it was in time. Partials that have not risen, such as a note already held under the attack, keep their carried phases and are not affected.

Fully left, at SMOOTH, nothing is marked, which suits pads and held sounds. Turning right lowers the threshold, from a rise of 24 dB near the left end to 3 dB fully right, at SHARP. It starts at its centre, 13.5 dB. No new attack is marked during the four frames that follow one. Its tooltip reads `Transient: a rise of 14 dB resets the phases.`

TRANSIENT does most for sharp, wide-band attacks at moderate shifts. Shifted +5 semitones, an impulse keeps 93% of its energy within a millisecond of where it was at SHARP, against 12% at SMOOTH. At large shifts it does less: at +12, 69% against 57%. A soft attack, such as a click lasting a millisecond, is not spread in the first place, and TRANSIENT does not change it.

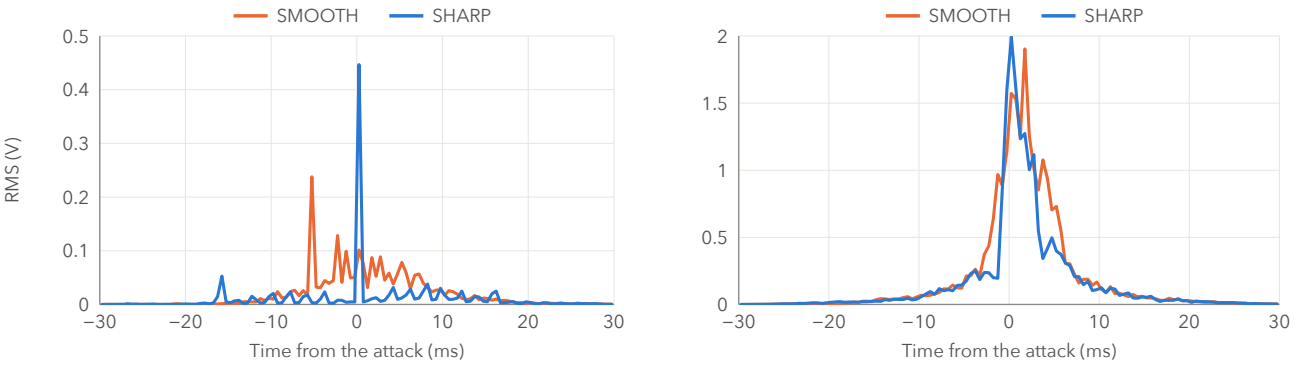


Figure 4. Left: a 3 V impulse shifted up 5 semitones. Right: a burst of noise decaying in 3 ms, like a closed hi-hat, shifted down 7 semitones. Each after silence, at 2K, its RMS at OUT L in 0.5 ms steps; 0 ms is where the attack was played. At SMOOTH the impulse is spread over the window, much of it ahead; at SHARP most of it stays at 0 ms. The burst at SHARP rises more sharply and has less ahead of it: 57% of its energy is within 1 ms of 0, against 33%.

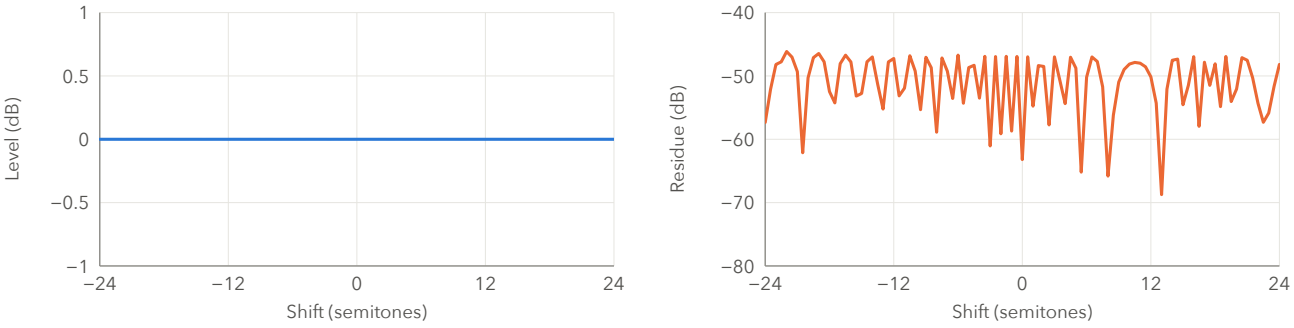


Figure 5. A 2 V sine at 440 Hz, shifted from −24 to +24 semitones in half-semitone steps, at 2K. Left: the level of the shifted sine, against the input's; it is within 0.001 dB at every shift. Right: everything else in the output, against the input's level.

2 · FORMANT



1 **FORMANT**	−12 to +12 semitones	2 **AMT**	FORMANT's CV amount
3 **CV**	a CV on FORMANT	4 **DETAIL**	the envelope's smoothing
5 **PRESERVE**	holds the envelope		

A sound's spectral envelope is the outline of its spectrum: how loud it is at each frequency, regardless of where its partials fall. In a voice the envelope's peaks are the formants, set by the shape of the throat and mouth, and they stay where they are whatever note is sung. Shifting every partial by the same ratio moves the envelope with them, and a voice shifted up sounds smaller.

Shiftr draws the input's envelope in each frame through the levels of its partials: it joins the peak of each partial to the next in decibels, then smooths the result across frequency by **DETAIL**. Drawn this way the envelope does not depend on how far apart the partials are, so it is the same shape for a low voice as for a high one.

Each shifted partial is then raised or lowered so that it sits on the target envelope at its new frequency. The correction is limited to 24 dB either way.

PRESERVE, when lit, keeps the envelope where it was: the target is the input's envelope at the same frequency. When it is off, where it starts, the envelope moves with the pitch, as in a plain pitch shift (Figure 6).

FORMANT moves the target envelope by up to 12 semitones either way, from where PRESERVE puts it. It starts at its centre, 0. With PRESERVE lit and PITCH at 0, FORMANT changes the envelope and not the pitch (Figure 7). With PRESERVE off, FORMANT moves the envelope further than the pitch moved it, or less far.

PRESERVE	FORMANT	The envelope
off	0	moves with the pitch: a plain shift; nothing is corrected
lit	0	stays where it was
lit	+4	moves up 4 semitones, whatever the pitch
off	−5, with PITCH +5	stays where it was, as PRESERVE does

FORMANT's CV, through its **AMT**, moves the envelope by 12 semitones for 5 V at full AMT, inverted to the left of AMT's centre. FORMANT and its CV together are held to ±12 semitones. Its tooltip reads `Formant: +4.0 semitones from where it was` with PRESERVE lit, and `... from where the pitch takes it` with PRESERVE off.

DETAIL sets how much the envelope is smoothed across frequency: by about 1 kHz fully left, at COARSE, and not at all fully right, at FINE. It starts at 0.6, a smoothing of 160 Hz. Smoothing makes the envelope steadier on noisy and breathy input, where every bin can be a peak; at FINE the envelope follows every partial.

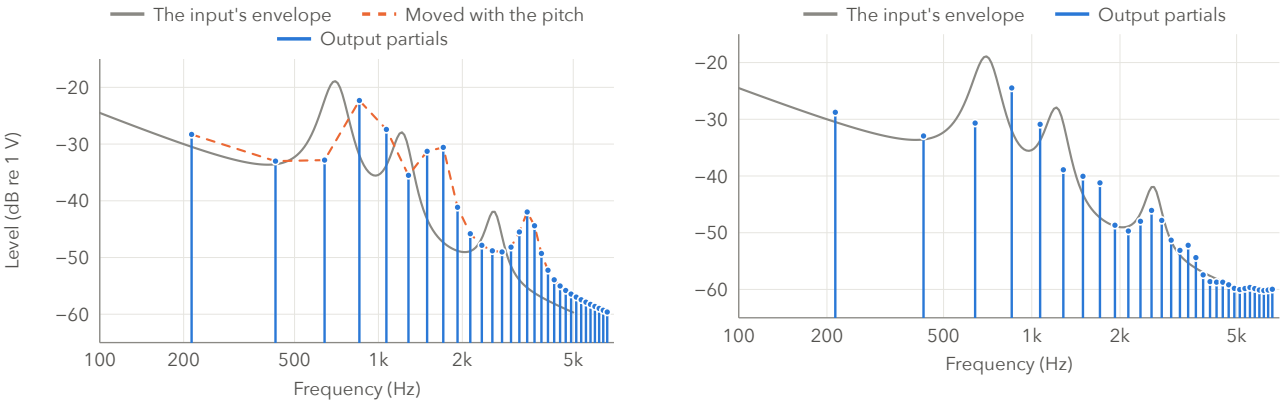


Figure 6. A vowel sung at 160 Hz, shifted up 5 semitones, at 2K. Left, PRESERVE off: the output's partials follow the envelope moved up with them. Right, PRESERVE lit: they follow the input's envelope where it was. Each stem is one partial of the output, measured at OUT L.

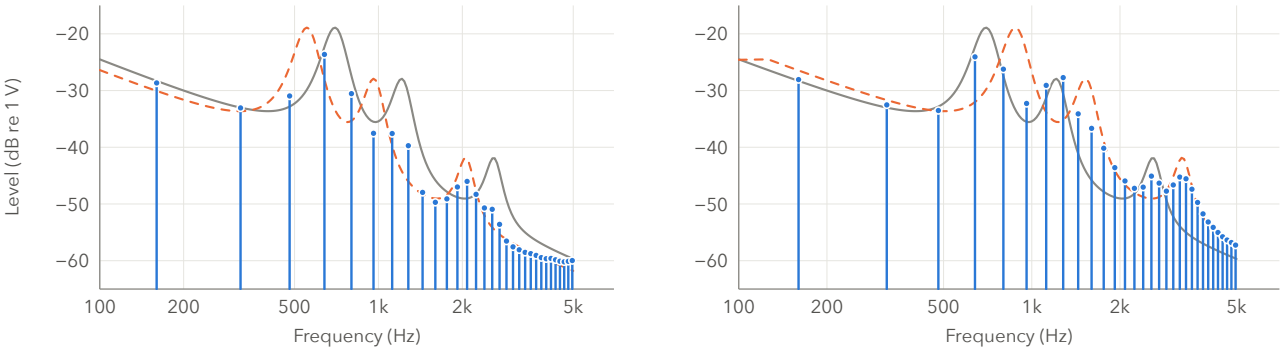


Figure 7. The same vowel at its own pitch, PRESERVE lit, FORMANT at −4 (left) and +4 (right). The partials stay at 160 Hz and its multiples; their levels follow the input's envelope (grey) moved down or up 4 semitones (dashed).

3 · FEEDBACK and FREEZE



1	**FEEDBACK**	0 to 98% back to the input	2	**AMT**	FEEDBACK's CV amount
3	**CV**	a CV on FEEDBACK	4	**DAMP**	the loop's low-pass
5	**MODE**	HOLD, BLUR, DRIFT	6	**FREEZE**	holds the spectrum
7	**GATE**	FREEZE, while above 1 V			

FEEDBACK adds the shifted output back to the input, so that each pass through Shiftr is shifted again. With PITCH at +12, a note returns an octave higher, then two octaves higher, and so on, each return one window and a quarter after the last: every 53.3 ms at 2K (Figure 8). At +7 the returns rise in fifths. Each return is higher than the last, so a note's returns end when they pass 24 kHz: a 110 Hz note at +12 makes six, at 440 Hz to 14 kHz. At PITCH 0, FEEDBACK repeats the input every window and a quarter, as a delay does. It starts fully left, off. Fully right, 98% of the output is returned. Its tooltip gives the percentage returned.

FEEDBACK's CV, through its **AMT**, moves FEEDBACK across its whole range for 5 V at full AMT. The two together are held between 0 and 98%.

The returned signal passes through a soft limit, the same as the output's (see Section 4), then **DAMP**, then a high-pass filter at about 4 Hz that removes any DC. The loop's gain is below one at every setting, and the limit holds the returned signal below 10 V, so feedback cannot run away.

DAMP is a low-pass filter in the loop, 6 dB per octave. Its corner is 20 kHz fully left and 1 kHz fully right, equally spaced in ratio between. It starts at 0.3, 8.1 kHz. Each return passes through it once more, so the higher returns, which have passed through it more times, are quieter. Its tooltip reads `Damp: the loop rolls off above 8142 Hz.`

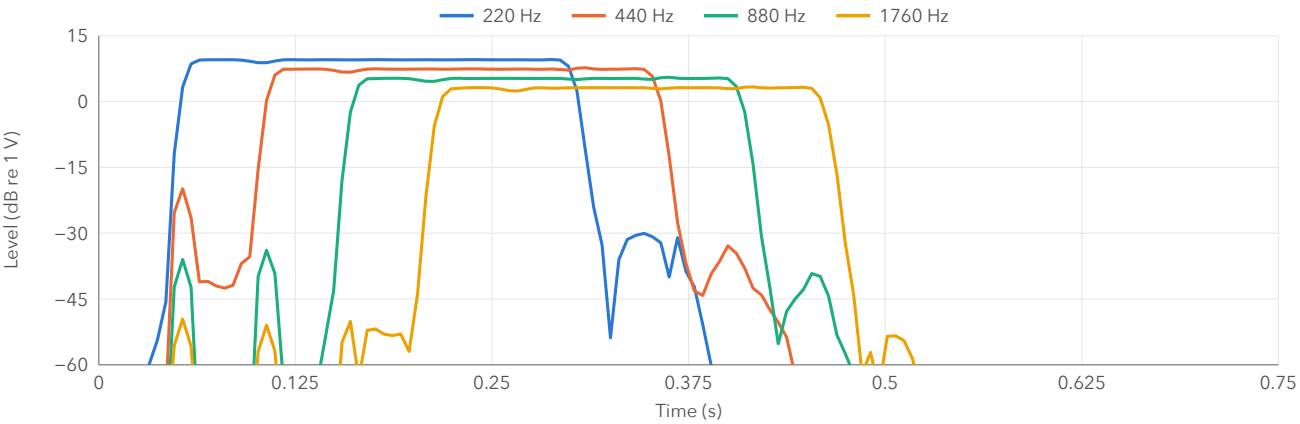


Figure 8. PITCH +12, FEEDBACK 0.8 (78% returned), DAMP fully left, at 2K: a quarter second of 110 Hz at 3 V, then silence. The first pass is at 220 Hz; each return is an octave higher, about 2 dB quieter, and arrives 53.3 ms after the one before. Returns above 1760 Hz are not plotted. Measured at OUT L in 21 ms windows.

FREEZE, when lit, holds the spectrum and keeps it sounding until FREEZE is released; the input is ignored meanwhile. PITCH, FINE, PITCH MOD, FORMANT, PRESERVE, SPREAD and MIX still act on the held spectrum, so turning PITCH

while frozen plays the held sound at other pitches. FEEDBACK and DAMP have no effect on it. When FREEZE is released, Shiftr returns to its input at the next frame.

MODE sets what FREEZE holds:

MODE	What is held
HOLD	The last frame analysed, exactly. The same frame is played again in every frame interval, so everything held that is not a steady partial, noise above all, is heard as lines spaced at the frame rate: 94 Hz apart at 2K, 188 Hz at 1K, 47 Hz at 4K. They sound as a buzz at that pitch (Figure 9).
BLUR	The input's power averaged over about the last eight frames (85 ms at 2K). Peaks that are steady partials are held as steady tones. Peaks that are noise take new random phases in every frame, so the noise is held as noise of the same level and colour. It starts here.
DRIFT	As BLUR, and each held partial moves slowly up and down by up to 6 cents, at its own rate between one cycle in 14 s and one in 4 s, so that a held chord changes continuously.

Shiftr tells a partial from noise by its neighbouring bins: the bins either side of a partial's peak give the same frequency as the peak, to within 0.3 of a bin, and the bins beside a noise peak do not. The average BLUR and DRIFT hold is taken when FREEZE is pressed: changing MODE from HOLD while frozen changes how the held peaks are played, not which frame is held. Its tooltip names the mode, as `Freeze mode: BLUR, the recent average, noise kept noisy.`

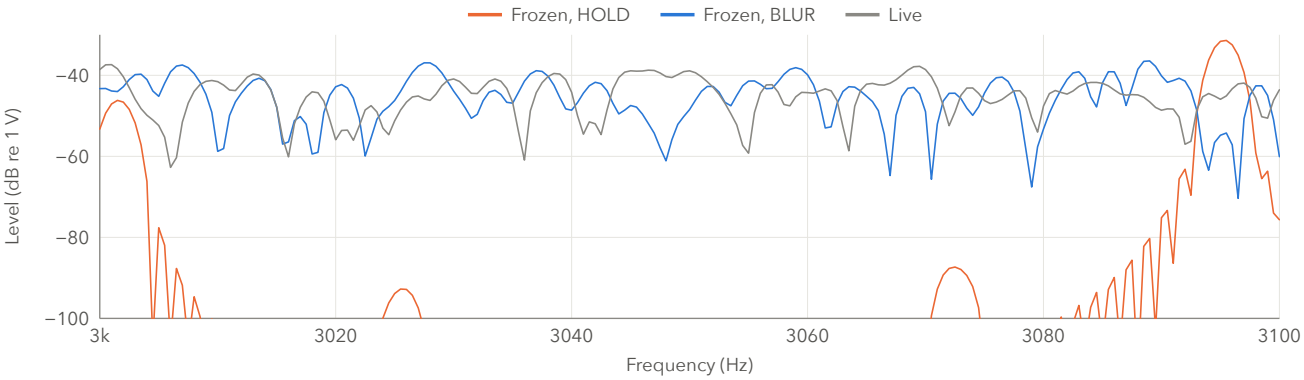


Figure 9. A 220 Hz tone over white noise, frozen at 2K: a 100 Hz stretch of the spectrum where there is only noise, at 1.5 Hz resolution. Live and in BLUR it is the irregular, continuous spectrum of noise, at the same level. In HOLD it is two lines 94 Hz apart, the frame rate at 2K, and almost nothing between them.

GATE freezes Shiftr while its voltage is high: it turns on above 1 V and off again below 0.5 V. The button and the gate are combined: Shiftr is frozen while either is on. The gate is read every sample and takes effect at the next frame.

4 · I/O



1	**IN L**	the left input	2	**IN R**	the right; IN L if unpatched
3	**MIX**	dry against shifted	4	**AMT**	MIX's CV amount
5	**CV**	a CV on MIX	6	**SPREAD**	0 to 50 cents between sides
7	**OUT L**	the left output	8	**OUT R**	the right output

IN L and **IN R** are the two channels' inputs. When nothing is patched into IN R, IN L feeds both channels.

MIX sets the dry signal against the shifted signal, with equal-power gains (Figure 11). It starts fully right: shifted only. The dry signal is the input delayed by the same time as the shifted signal (see Latency), so the two are aligned. With PITCH at 0 and nothing else changed, the two are the same signal, and MIX at its centre is 3 dB louder than at either end. MIX's CV, through its **AMT**, moves MIX across its whole range for 5 V at full AMT, and is read every sample. Its tooltip reads `Mix: 100% wet`.

SPREAD lowers the left channel's pitch and raises the right channel's, by up to 25 cents each: up to 50 cents between them, fully right. It starts fully left, where both channels are the same. It is added after PITCH, FINE and PITCH MOD are held to ±24 semitones. With PRESERVE off, the envelope follows each channel's own pitch.

OUT L and **OUT R** are the two channels' outputs, after MIX. Each passes through a soft limit that leaves signals between −5 V and +5 V unchanged, and above 5 V reduces them progressively so that the output never exceeds 10 V and never folds back (Figure 10).

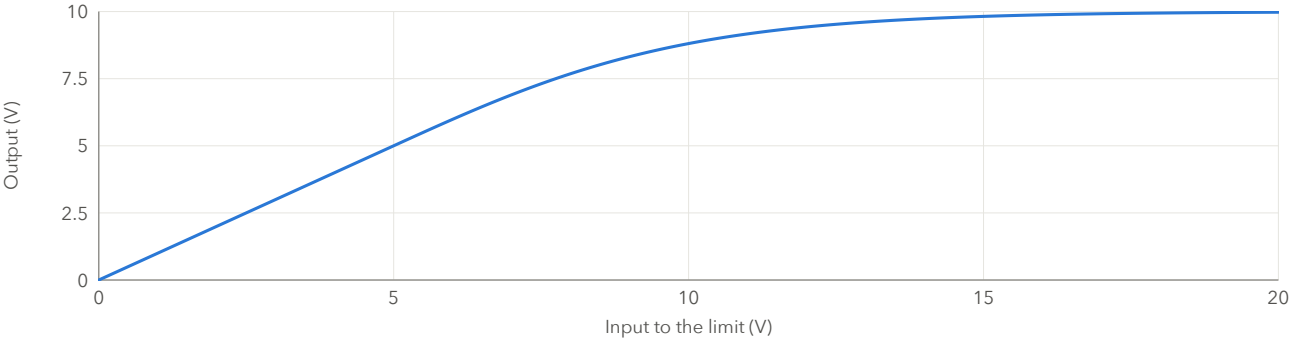


Figure 10. The output's soft limit, the same for negative voltages. It is unity to 5 V.

5 · The display

The panel in the header shows the left channel's most recent frame, from 60 Hz to 12 kHz on a logarithmic frequency axis, divided into 64 bands, and from −66 dB to 0 dB re 5 V. In each band:

- a grey line is the loudest input partial in that band;
- a pink line beside it is the loudest shifted partial in that band;
- a white line across the bands is the target envelope that the shifted partials are fitted to.

With PRESERVE lit and FORMANT at 0, the pink lines move as PITCH changes and the white line stays where it is. With PRESERVE off, the white line moves with the pitch. A line rises at once and falls by 1.5 dB each time the display is redrawn, about 30 dB a second. The display has no controls and does not affect the sound.

6 · How the shift works

Each channel cuts its input into frames of WINDOW samples, starting every quarter of a window, and multiplies each by a Hann window. For each frame it computes the spectrum with a fast Fourier transform, then:

1. **Finds the peaks.** A peak is a bin louder than both its neighbours. Each peak owns the bins out to the quietest bin between it and the next peak; that group of bins is the peak's region.
2. **Measures each peak's frequency.** The phase of the peak's bin advances from one frame to the next at a rate set by the partial's exact frequency. Shiftr measures that advance, and so finds the frequency to a small fraction of a bin.
3. **Draws the envelope** through the peaks' levels, as Section 2 describes.
4. **Moves each region** by the difference between the peak's frequency and the shifted frequency, which is generally a fraction of a bin. The region's shape and level are kept: a sine shifted by any amount has the input's level to within 0.001 dB.
5. **Turns each region's phases** by one angle, carried on from the region's angle in the previous frame by the change in frequency over one frame interval. All the bins of a region turn together, so the partial's shape is kept. At a shift of 0 the angle stays at 0 and the output is the input, delayed.
6. **Scales each region** to the target envelope at its new frequency (Section 2), if PRESERVE is lit or FORMANT is away from 0.

The frame is then turned back into a signal by the inverse transform, multiplied by the window again, and added to the frames on either side of it.

The work of each frame is spread over the samples of its frame interval, on the audio thread, so that no single sample carries a whole frame's work.

What the method does to a sound

- **Attacks are spread.** Each frame covers one window, so a sudden attack can be heard spread over up to one window: 21 ms at 1K, 43 ms at 2K, 85 ms at 4K. TRANSIENT keeps most attacks in place (Section 1); shorter windows spread what is left less.
- **Low notes need long windows.** Partials closer together than about two bins cannot be separated (see Latency). At 1K that is about 94 Hz, so a note below about 90 Hz has partials closer together than that, and is shifted as a blur. At 4K the limit is about 23 Hz. Use 4K for bass and chords.
- **Noise is shifted as partials.** Noise has a peak every few bins. Each is moved as a partial, which keeps the noise's level and moves its colour with the pitch.
- **Large shifts down lose the top.** Shifted down an octave, nothing comes out above 12 kHz, because the input has nothing above 24 kHz to bring down. Shifted up, partials that would land above 24 kHz are dropped.

7 · Control ranges

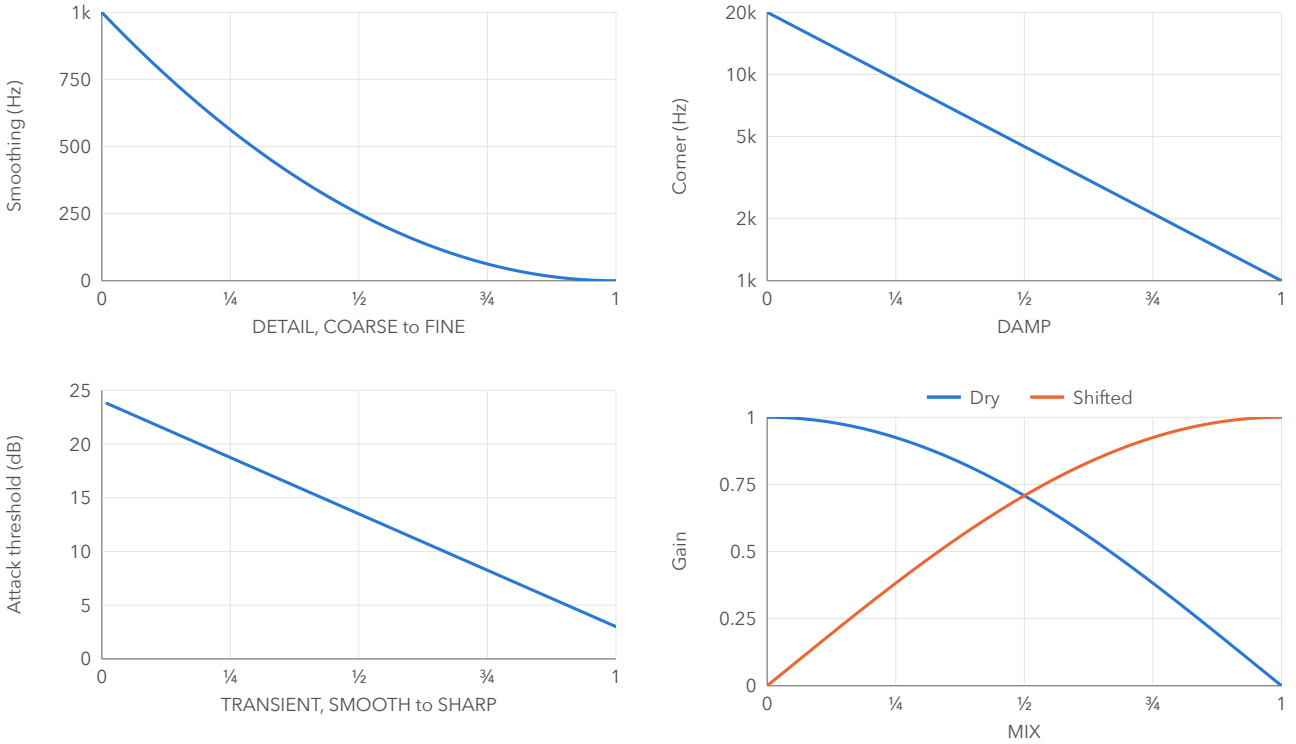


Figure 11. The knob laws. Each is the function the engine runs. TRANSIENT has no threshold fully left, where no attack is marked.

Quantity	Value	Notes
Shift	−24 to +24 semitones	PITCH, FINE and PITCH MOD together
PITCH MOD	12 × AMT × |AMT| semitones/V	1 V/oct at either end of AMT
Formant	−12 to +12 semitones	FORMANT and its CV together; 12 semitones per 5 V
Envelope correction	±24 dB at most	per partial
Envelope smoothing	1 kHz to none	DETAIL; 160 Hz where it starts
Window	1024, 2048, 4096 samples	Hann; a new frame every quarter window
Delay	1.25 windows	26.7, 53.3 or 106.7 ms; the dry is delayed the same
Attack threshold	24 to 3 dB	TRANSIENT; none fully left; at most one attack in four frames
Feedback	0 to 98%	FEEDBACK and its CV together
DAMP	20 kHz to 1 kHz	6 dB/oct; 8.1 kHz where it starts
Feedback DC block	about 4 Hz	
SPREAD	0 to 50 cents	left down and right up, by half each
FREEZE GATE	on above 1 V, off below 0.5 V	
FREEZE modes	HOLD, BLUR, DRIFT	BLUR: about 8 frames' average; DRIFT: ±6 cents, 0.07 to 0.25 Hz
Output	unity to ±5 V	soft limit, never above 10 V
When CVs take effect	PITCH MOD, FORMANT: each frame	MIX, FEEDBACK: each sample

Jack	Direction	What it carries

PITCH MOD	in	the shift; up to 1 V/oct, by AMT
FORMANT CV	in	the envelope's position; 12 semitones per 5 V at full AMT
FEEDBACK CV	in	FEEDBACK; its whole range per 5 V at full AMT
GATE	in	FREEZE, while above 1 V
IN L, **IN R**	in	the audio; IN R is IN L when unpatched
MIX CV	in	MIX; its whole range per 5 V at full AMT
OUT L, **OUT R**	out	the audio, after MIX and the soft limit

Shiftr saves nothing but its controls. A frozen spectrum and the feedback loop's contents are not saved with a patch.

The measurements in this manual are taken from the module's code, running its engine at 48 kHz. The test vowel is a sum of partials of 160 Hz up to 5 kHz, each at the level of a vowel envelope with formants at 700, 1220 and 2600 Hz.

Troubleshooting

- **There is no shift:** PITCH, FINE and PITCH MOD add to 0, or MIX is fully left.
- **The output is late against other parts:** Shiftr delays everything, dry included, by 1.25 windows (see Latency). Use 1K for the shortest delay.
- **A bass note sounds blurred or rough:** WINDOW is too short for it. Use 4K.
- **Attacks sound soft or early:** turn TRANSIENT towards SHARP. If they still do, use 1K.
- **PRESERVE seems to do nothing:** PITCH is 0, or the input has no pronounced envelope. The effect is strongest on voices.
- **A voice sounds hollow with PRESERVE lit:** DETAIL is too far towards FINE for a breathy or noisy voice. Turn it left.
- **A patched CV does nothing:** its AMT knob is at its centre.
- **Turning PITCH MOD's AMT a little does a lot:** AMT is not linear. The first quarter of its travel covers up to 0.75 semitones per volt.