We start with a system (in Laplace-land)

$$H(s) = \frac{b}{s+a}$$

but $H(s) = Y(s)/X(s)$, so

$$Y(s)(s+a) = bX(s)$$

As you hopefully remember from a linear systems class, 's' in the Laplace transform corresponds to a time derivative, so in the time domain we get

$$y'(t) + ay(t) = bx(t) \tag{1}$$

In general, it's true from calculus that

$$y(t) = y(t_0) + \int_{t_0}^{t} y'(\tau)d\tau$$

We approximate the integral by the trapezoidal rule, and evaluate at the points $t = nT$ and and $t_0 = (n-1)T$ (notice that by moving to time sampled at intervals of T, we are moving from continuous to sampled time). Then, the above becomes

$$y(nT) \approx y(nT-T) + T\left(\frac{y'(nT) + y'(nT-T)}{2}\right) \tag{2}$$

However, Eq. 1 gives a result for the derivative. We can plug this into Eq 2, and do a bunch of algebra. If we define $y(n) = y(nT)$ for simplicity, we have

$$(1 + aT/2)y(n) - (1 - aT/2)y(n-1) = bT/2(x(n) + x(n-1))$$

If we take the z-transform (as we are now in discrete time), we get

$$(1 + aT/2)Y(z) - (1 - aT/2)z^{-1}Y(z) = bT/2(1 + z^{-1})X(z)$$

Collecting $Y(z)$ terms and rearranging, we can find

$$H(z) = \frac{Y(z)}{X(z)} = \frac{b}{\frac{2}{T}\frac{1-z^{-1}}{1+z^{-1}} + a}$$

but

$$H(z) \approx H(s) = \frac{b}{s + a}$$

(the reason it's approximate is that the trapezoidal rule only approximated the integral). Thus

$$s \approx \left(\frac{2}{T}\right)\frac{1 - z^{-1}}{1 + z^{-1}}$$

which is the bilinear transform.

Now, we can write s in Cartesian coordinates as $s = \sigma + j\Omega$ and z in polar coordinates as $z = re^{j\omega}$. Plugging into the equation above and doing some algebra, we get an equation for frequency warp:

$$\Omega = \frac{2}{T}\tan(\omega/2)$$

which can also be written as

$$\omega = 2\arctan\left(\frac{\Omega T}{2}\right)$$