Appendix 1

Wavelets

American Scientist 82 (April 1994) 250-255

Gilbert Strang

I was listening to Mozart, but my mind was wandering. Normally one doesn't admit such a thing. Possibly this was not Mozart at his best. Symphony Hall was full, we were late, and parking had been very illegal. But the real reason for inattention was an idle thought, and it led me toward this article. The thought was really a question: *How did he write it all down?*

For Mozart that must have been a serious problem. Musical notation is fairly efficient for its purpose, but a full score for all instruments equals a small book. The *notation* is crucial. To a musician it may seem the only possible notation — but it's not. When that signal is recorded (say digitally), the same symphony is expressed in a totally different way. The score that Mozart wrote is not transmitted.

The complete signal is not only the sequence of notes, but the nuances that the performer adds. The touch on the keys or the pedal or the bow — these distinguish a master. There may be an extra word saying *forte* or *pianissimo*. This is not too precise! Mathematically, the notes are an indication but not a complete code.

Our world is full of sounds and images and bit strings that have to be compressed and transmitted and recovered:

- Audio — music and speech
- Video — still images and television
- Data — numbers, letters, books, even zip codes.

Zip codes seem trivial, but this short signal illustrates the choices to be made. I think that England and Canada overdid the compression. Strings like M5S 1A4 and V6T 1W5 give extra precision, but they are too hard to list and remember. (Does any reader knows how those countries chose alphanumeric?) Compression is delicate, because we have to *process* the signal, not just code it. It is the same for a student taking notes in class: compress the lecture but stay readable.

Television has been avoiding compression, but that is about to change. Each "pixel" on the TV screen needs 24 zeros or ones to mix three colors. The shading levels go from 0 to 255, making 2^8 possibilities — eight bits for each color. A very sharp image needs more pixels with more bits than we can afford to send or receive. Compression is required; information has to be thrown away. Otherwise the TV can't keep up with real time!

The problem is really severe for *high definition television*. It has an enormous bit rate. We will soon see an amazingly clear picture with many more pixels, provided the compression is well done. The processing of HDTV signals is a billion-dollar decision that the biggest companies are competing for. At the end I will mention a personal experience with this mathematical and engineering and increasingly political competition.

Fourier Transform and Wavelet Transform

This article is about several methods — none perfect, all being improved — to analyze and synthesize a signal. I mention three ways to transform a symphony:

1. Into cosine waves (by Fourier transform)
2. Into pieces of cosines (short time Fourier transform)
3. Into wavelets (the new way).

The classical method is to separate the whole symphony into pure harmonics. Each instrument plays one steady note. The signal becomes a sum $a_0 + a_1 \cos t + \ldots$ of cosine waves. This is the *Fourier transform*, published 172 years ago by Joseph Fourier in Paris. He assumed an infinite orchestra: Conductor not needed, musicians totally bored.

All music and all signals can be analyzed into pure harmonic waves by the Fourier transform. Engineers do that almost instinctively — the signal strength at each moment in time is replaced by the amplitude of each wave. The big question is how many frequencies are needed for a high-fidelity signal. Probably too many for good compression.

A second possibility is the *Short Time Fourier Transform.* Short segments of the symphony are transformed separately. In each segment, the signal is separated into cosine waves as before. The musicians still play one note each, but they change amplitude in each segment. This is the way most long signals are carved up.

One disadvantage: There are sudden breaks between segments. You might or might not hear that. In a visual image you could probably see it as an edge. This "blocking effect" is the nemesis of the Short Time Fourier Transform.

A new idea in signal processing is closer to the score that Mozart originally wrote. Instead of cosine waves that go on forever or get chopped off, the new building blocks are "*wavelets*". These are little waves that start and stop. They all come from one basic wavelet $w(t)$, which gives the sound level of the standard tune at time t. The basses spread that tune over the longest time. The cellos play the same tune but in half the time, at doubled frequency. Mathematically, this speed-up replaces the time variable t by $2t$. The first bass plays $b_1 w(t)$ and the first cello plays $c_1 w(2t)$, both starting at time $t = 0$.

The next bass and cello play $b_2 w(t-1)$ and $c_2 w(2t-1)$, with amplitudes b_2 and c_2. The bass starts when $t - 1 = 0$, at time $t = 1$. The cello starts earlier, when $2t - 1 = 0$ and $t = \frac{1}{2}$. There are twice as many cellos as basses, to fill the total length of the symphony. Violas and violins play the same passage but faster and faster and all overlapping. At every level the frequency is doubled (up one octave) and there is new richness of detail. "Hyperviolins" are playing at 16 and 32 and 64 times the bass frequency, probably with very small amplitudes. Somehow these wavelets add up to a complete symphony. This article will attempt to show how.

Summary: A long signal is broken into a *basis of possible signals.* Those can be notes or waves or wavelets. The wavelets come from a single function $w(t)$ by speed-ups and time delays. Amplitudes are sent to the receiver, which reconstructs the symphony. (Very small amplitudes are discarded; more later about this crucial compression step.) We now explain the idea of a basis.

Choosing the Best Basis

Some readers will know about vectors and matrices. This section is for you. You remember that every vector (x, y) is a combination of the basis vectors $(1, 0)$ and $(0, 1)$. The vector $(4, 2)$ is 4 times the first basis vector plus 2 times the second.

Another pair of basis vectors is $(1,1)$ and $(1,-1)$. Again they produce all plane vectors, including $(4,2)$. The requirement met by both bases is this: *Each vector in the space can be expressed in one way only as a combination of the basis vectors.* The vectors $(1,0)$ and $(2,0)$ would not be a basis. They lie on the same line; no combination can give $(4,2)$. The vectors $(1,0)$ and $(0,1)$ and $(3,1)$ are not a basis — three vectors is too many. The signal $(4,2)$ can be produced from any two of them (and it is also the sum of all three). Any two vectors in different directions, like $(5,1)$ and $(3,1)$, form a basis for the whole plane.

The best bases have a valuable extra property: The vectors are perpendicular. For the standard basis $(1,0)$ and $(0,1)$, across is perpendicular to up. The vectors $(1,1)$ and $(1,-1)$ also pass the test for a $90°$ angle: *their dot product is zero.* The dot product of (x,y) with (X,Y) is $xX+yY$. The dot product of $(1,1)$ with $(1,-1)$ is $1-1=0$. The dot product of $(5,1)$ and $(3,1)$ is 16 — not perpendicular. More to the point, a vector with 1000 components separates quickly into 1000 perpendicular pieces — speed is crucial. These steps are at the heart of linear algebra.

For engineers and social and physical scientists, *linear algebra* now fills a place that is often more important than calculus. My generation of students, and certainly my teachers, did not see this change coming. It is partly the move from analog to digital; functions are replaced by vectors. Linear algebra combines the insight of n-dimensional space with the applications of matrices.

Higher Dimensions

Move now to four dimensions. Think of the components (x_1,x_2,x_3,x_4) as strengths of a signal. If the signal is $x=(1,1,1,1)$ you are producing a steady note. With $x=(1,2,4,8)$ the volume is increasing.

The easy way to send the signal is to give its four components. Implicitly, you have chosen the standard basis in four-dimensional space. The basis vectors are $(1,0,0,0)$ and $(0,1,0,0)$ and $(0,0,1,0)$ and $(0,0,0,1)$. You are sending the numbers x_1,x_2,x_3,x_4 that multiply your basis vectors. How could this be less than optimal? The answer depends on the signal.

Suppose you hold the same note. This vector $(1,1,1,1)$ is very possible and very common, especially if the time interval is short. Good sound reproduction will certainly use short intervals. With the standard basis, you have to send 1 and 1 and 1 and 1. It would be better if the useful vector $(1,1,1,1)$ was actually in the basis, so one signal would do. *A good basis allows accurate reproduction with only a few vectors.*

Compression discards basis vectors that are absent (or barely present) in the signal. When the vectors represent different frequencies — the *Fourier basis* — high frequencies can often be discarded. A "lowpass filter" removes them. With the standard basis, we can discard $(0,1,0,0)$ only when the second note is not played. Otherwise there is a very audible (or inaudible) gap in the broadcast.

Musical notation has met this problem. It represents $(1,1,1,1)$ by a full note and also by two half-notes. The same signal has several forms. This isn't a basis! Full notes and half-notes and quarter-notes are fine for the musician, but the mathematician only wants four vectors in the basis.

It is quite pleasant to devise a basis that includes the constant vector $(1,1,1,1)$. The second basis vector can be $(1,1,-1,-1)$. This is a square wavelet, once up and once down. The German mathematician Alfred Haar completed a perpendicular basis in 1910, by *dilation* (or squeezing) and *translation* (or shifting). He squeezed $(1,1,-1,-1)$ to produce the third vector $(1,-1,0,0)$. Then he shifted by two time intervals to produce the fourth vector $(0,0,1,-1)$. The basis vectors go into the columns of the four-by-four "Haar matrix":

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & -1 & 0 \\ 1 & -1 & 0 & 1 \\ 1 & -1 & 0 & -1 \end{bmatrix}.$$

Can you guess the matrix H_8? Its first column is the constant vector $(1,1,1,1,1,1,1,1)$. Electrical engineers would call this the zeroth column; they start from zero, like English lifts. The next column is $(1,1,1,1,-1,-1,-1,-1)$, the wavelet that is played by the bass. It is squeezed to $(1,1,-1,-1,0,0,0,0)$ for the first cello, and shifted for the second cello. The last four columns of H_8 are viola notes like $(1,-1,0,0,0,0,0,0)$, with higher frequency over a shorter time. All these vectors are perpendicular!

The violins enter when the dimension reaches 16. I don't recommend writing down the matrix H_{16}, but you would know how. This is typical of linear algebra: When four dimensions are understood, higher dimensions are no problem. There are four violas and eight violins. The count $1+2+4+8$ equals 15, one short. The constant vector $(1,1,\ldots,1,1)$ makes the 16th or zeroth basis vector, which is the scaling function.

Fast Wavelet Transform

The very smallest wavelet transform we saw already. $(4,2)$ is 3 times $(1,1)$ plus 1 times $(1,-1)$. From the amplitudes 3 and 1, we recover the vector $(4,2)$. To find those amplitudes for longer signals, we use *averages* and *differences*. The averages move up in a pyramid, to be averaged again. The differences give the fine details at each level. When a signal changes quickly, those details are important. When the signal is almost steady, we might not transmit them. That is how compression works, by ignoring what the eye can't see and the ear can't hear.

The test case is in four dimensions, for a vector like (4,2,5,5). We know how 4,2 in the first half yields 3 and 1. The second half 5,5 yields 5 (the average) and 0 (the difference). The two averages 3 and 5 move up the pyramid. The two differences 1 and 0 are the details that go with the basis vectors $(1,-1,0,0)$ and $(0,0,1,-1)$.

The averages give a coarse signal $(3,5)$. Treat that the usual way. Compute $\frac{1}{2}(3+5)=4$ and the difference $\frac{1}{2}(3-5)=-1$. Consistency is the key to a good algorithm! The wavelet transform of $(4,2,5,5)$ is the set of averages and differences $(4,-1,1,0)$. This is transmited to the receiver, which does an inverse transform back to $(4,2,5,5)$. The decoder goes down the pyramid, first using the coarse parts 4 and -1 to recover 3 and 5. Those are combined with the fine details 1 and 0 to produce $4,2$ and $5,5$. We have reconstructed the signal.

This gives hope for a signal with 256 numbers. The pyramid has 8 levels because $256=2^8$. the lowest level starts as always with $A=\frac{1}{2}(x_1+x_2)$ and $D=\frac{1}{2}(x_1-x_2)$. The average A moves up to the next level. The difference D is the detail at this finest level.

The key word behind this pyramid is **multiresolution**. We are seeing the signal at multiple scales — fine, medium, and coarse. We resolve it at each scale by details and averages. Possibly our eyes do the same, to see the target and then the bull's-eye. Certainly our ears are designed to pick off higher frequencies as the sound travels into the cochlea. The art of the engineer is imitating nature.

Note: We called this wavelet transform "*fast*". A vector of N numbers is to be expressed as a combination of N basis vectors. In matrix language, we are multiplying by the Haar matrix H

or its inverse. Normally such a multiplication takes N^2 steps. The pyramid does it with only N averages and N differences.

Question: Is there also a fast Fourier transform? Can we quickly find the amplitudes of N cosine waves? Instead of H we use the Fourier matrix F, to transform between "time domain" and "frequency domain". This fast transform is the most important numerical algorithm in our lifetime, a very tough competitor for wavelets. It is wired into computers and explained in linear algebra textbooks. The key is to produce F out of matrices whose entries are almost all zeros. The heart of numerical linear algebra has become matrix factorizations — H and F are outstanding examples.

The Daubechies Wavelets

Wavelets are a hot topic, but the Haar wavelets are cold. They were invented in 1910. Their graphs are made from flat pieces; anything that simple has already been thought of. Approximation to most signals is very poor. We need many many flat pieces to represent even a sloping line to decent accuracy. The basis will not give compression ratios of 20:1 or 100:1, as desired. So we turn to a better basis.

The new wavelets are more intricate, and their only formula is an infinite product, but eventually mathematics had to find them. I will touch on this advance by describing the discovery made by Ingrid Daubechies. In 1988, at AT&T Bell Laboratories, she found a pulse that starts and stops and is perpendicular to all its dilations and shifts. It is based on four "magic" numbers h_0, h_1, h_2, h_3. Her scaling vector S uses them in that order. Her wavelet uses them in the order $W = (h_3, -h_2, h_1, -h_0)$. Do you see why W is perpendicular to S? Multiply and add: the dot product $S \cdot W$ cancels itself to give zero. She also wanted $(1,1,1,1)$ and $(1,2,3,4)$ to have zero component along W, so constant and linear signals can be greatly compressed. Their dot products with W must be zero:

$$h_3 - h_2 + h_1 - h_0 = 0 \quad \text{and} \quad h_3 - 2h_2 + 3h_1 - 4h_0 = 0.$$

Those are two equations for the h's; we need two more. The third equation makes the first bass tune $(h_3, -h_2, h_1, -h_0, 0, 0)$ perpendicular to the second bass tune $(0, 0, h_3, -h_2, h_1, -h_0)$. Their dot product is required to be $h_1h_3 + h_0h_2 = 0$. Then a fourth equation $h_0 + h_1 + h_2 + h_3 = 2$ sets the size of the h's. Her numbers from solving the four equations give a much better filter than Haar: $4h_0 = 1 + \sqrt{3}$, $4h_1 = 3 + \sqrt{3}$, $4h_2 = 3 - \sqrt{3}$, $4h_3 = 1 - \sqrt{3}$. I must mention that this is not the ultimate. Six numbers or eight numbers can be even better, but also more work. Video filters tend to be short. Audio filters are amazingly long, because music is generally smoother than a picture.

The key step from discrete vectors to continuous functions is the *dilation equation.* This uses the magic numbers h_0, h_1, h_2, h_3 in a curious but natural way. The equation for the scaling function $\phi(t)$ involves t and $2t$ and the magic h's:

$$\phi(t) = h_0\phi(2t) + h_1\phi(2t-1) + h_2\phi(2t-2) + h_3\phi(2t-3).$$

Substitute $t = 1$ and $t = 2$ to find $\phi(1)$ and $\phi(2)$. Then the equation gives ϕ at $t = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$ because $2t$ on the right side is a whole number. From these half-integer times we go to quarter-integers. Eventually we have enough values to draw the graph from $t = 0$ to $t = 3$. This has become famous (to some people) as a function with entirely new properties, all built into the dilation equation.

Where Fourier used a cosine wave and Haar used a square wave, Daubechies begins with this scaling function. Her wavelet $w(t)$ has the same right side as the equation above, but with coefficients $h_3, -h_2, h_1, -h_0$. It has a highly irregular graph, a type of fractal. Squeezing and shifting it gives the complete wavelet basis. But all computations go back to the four numbers.

High Definition Television

"The shot heard round the world" was a metaphor in 1775. The signal seen around the world is now a reality. It is television. You may think that the signal could be improved (by better programs). Engineers also want to improve it (by more pixels and better compression). Probably this second improvement is the one we will get, with HDTV.

The battle to set the standard for high definition television has been played out in a surprisingly public way. Europe invested a billion dollars, then stopped. It was decided to follow the North American standard. A competition was organized by the Federal Communications Commission, and there were four finalists. One big issue is *motion estimation*, to predict where the picture will move. Then you only have to transmit a small correction, and you can keep up with all the pixels in real time.

The New York Times expected that the Japanese entry would win, because HDTV is already available in Japan. But it is based on analog signals, not 0's and 1's. Digital filters are now well developed, to separate the high frequencies and compress that information. But sharp edges have to stay sharp. Experts know exactly what to look for, like tasting wine or tea.

To the great credit of the New York Times, the competition was explained to the public. Every month it took a new turn. The FCC finished its tests, but one group of companies wanted a makeup test. They had stayed up the night before and their system wasn't feeling too well. (All professors recognize these symptoms, but the government is softer. They allowed a retest.) We were expecting one winner, but the FCC urged cooperation in the end. We all hope it works.

I must explain how I learned about this. As usual, it was from a student. He came into my office and said "Do you remember me?" Well, I tried to think fast. Too many students, too weak a memory, and better to be absolutely honest. "Not completely," I replied — never expecting to hear what he told me next. "You are my thesis advisor."

What do you say to a thesis student you don't remember? In that position I suggest something very short: "Tell me more." The most amazing part was his thesis topic. "I am designing the filter bank for MIT's entry in the HDTV competition." Some days you can't lose, even if you deserve to.

It is true that I am his advisor. He is a mathematics student who looked for a thesis in electrical engineering. He found Jae Lim, who masterminded MIT's joint entry with General Instruments. A formal committee was required by the mathematics department. I must have agreed, since my signature was there on the paper. But my very own student would not tell me the coefficients $h_0, \ldots, h_6$ in his filters. He expressed confidence in his advisor, but not enough for a billion-dollar secret.

Summary

A family of wavelets is an orthogonal basis. By combining the wavelets we represent any signal. Music has one independent variable: time. A photograph has two variables, across and up. Then wavelets in x multiply wavelets in y. A video image has three variables, x, y, and t — a flow of pictures with very important predictability. The competition between bases is fierce. Where computers race for faster calculations, mathematics races for quicker algorithms. An idea that cuts in half the *number* of steps is as good as a chip that doubles the speed. I don't know if the idea of wavelets is in that league, but it might be.

The HDTV standard is based on Fourier transforms. Wavelets came too late to have a real chance in that video race, but they are highly competitive for audio. Four stereo channels need only a small fraction of a TV band. In a completely different competition, run by the FBI and worth describing here, wavelets have won.

The FBI has 30 million sets of fingerprints, more every day. They need to be digitized. When a driver is stopped, or a thief is apprehended, fingerprints will be recorded electronically. A central computer will look for a match. As it is now, with thumb prints in enormous files in Washington, your chance to escape has been pretty good. Soon the "minutiae" of ridge endings and bifurcations will catch you.

It was expected that the older Fourier methods would succeed here too. But with 20:1 compression, the fingerprint ridges couldn't be followed. Lines were broken between one 8 by 8 square and the next. The short time Fourier transform introduced too much blocking. Wavelets gave a better picture. Tom Hopper at the FBI has chosen that new basis.

So when you get caught, you can blame it on wavelets.