Using personal computers for acoustic analysis in the voice laboratory

Prof. Sten Ternström
Dept of Speech, Music and Hearing, School of Computer Science and Communication
Kungliga Tekniska Högskolan, Lindstedtsvägen 24
SE-100 44 Stockholm, Sweden
stern@kth.se

Introduction

The personal computer has brought a revolution in enabling us to do lab work in acoustics and other science on the office desktop. In some ways, the cost of equipment is not nearly the problem that it was a decade or two ago. Still, we must bear in mind that today’s PC is a consumer device. As such, it is designed to meet requirements that are often very different from those that apply for scientific work.

In this document, which will be updated at irregular intervals, I try to collect various bits of computer-related information for people who run a voice laboratory. Comments and contributions are welcomed, and will be acknowledged.

To the best of my knowledge, the information is correct, but of course I cannot assume any responsibility for delay, damage or loss sustained from following recommendations presented here, which admittedly in some cases are opinionated.

Before you buy

Discuss the contents of this document with your technical aide and/or with the central computer support team at your institution.

Cheap to buy may not be cheap to run

A knowledgeable person can put together a very capable system, using consumer hardware and software that is affordable or even free. Given the much higher price of a turnkey specialist system, it can be very tempting to do-it-yourself. Remember, though, that with a system you assemble yourself, you also have the burdens of validating that your selected combination of components really works as intended, and of keeping it operational across incessant software updates and staff turnovers. For specialist systems, a serious supplier assumes this responsibility. The money you initially save on acquiring a system might easily be wasted later in time spent on troubleshooting and maintenance.

Use a file server and a database

In all but the smallest clinics and departments, set up a central file server so that recordings made in the studio can be accessed by all who need to do so. A few commercial systems can link up with your hospital or university database, such that voice recordings are automatically appended to the patient's journal. Other systems come with their own patient database, which is better than none, but not as good as interoperability with existing databases. Institutional and national databases are maintained by professionals, whose salaries probably come out of your overheads anyway.

Dedicate the main machine

Run your primary voice laboratory or clinic recording studio on a workstation that is not used for anything else. Dedicating a computer reduces costs and reduces downtime. You minimize the risks of malware incidents and of version conflicts that can result from updates to seemingly irrelevant parts of the system. Machines that are used in production should be kept free from all software that is not necessary for the job. During a trial period, you will probably install all manner of software and drivers, but once you have decided, reformat the hard disk, and reinstall the operating system along with the software that you will be using and nothing else. Do not install any e-mail, office applications or media players that are less than essential. If possible, consider disabling the computer's Internet connection in normal use, to minimize the risk of malware contamination.

Voice assessment on office computers

Some types of voice assessment do not require studio quality recordings. For example, measuring only the F0 distribution can be done in a quiet office, provided that basic acoustic requirements are met. If your site has several therapists, you might want to equip their office machines with a baseline inventory of recording/analysis software so that they can see patients in parallel. Do ensure that the storage of recordings is centralized, though.

User access control

This section applies mostly to computers running Microsoft Windows. If your Windows computers are centrally managed, the persons who will be operating the voice lab system will probably have restricted user rights on the operating system. If so, you must run a trial period to check that the lab software works properly when the user is logged in without administrator privileges.

Problems with strict access control are especially common when moving an old software application onto a newer operating system, as when upgrading Windows, for example. Even new software can suffer from insufficient testing, and problems occur even in products from large software houses. Do not accept the workaround to always log on as an administrator or 'power user', as this may compromise the security of your system.

Background: most voice analysis software has a university lab heritage; it was originally written by people who are researchers first and programmers second. Chances are that the software was written before Microsoft tightened up the access control systems in Windows, or that the author for convenience always was logged in as administrator when writing and testing. The programs presumably work on the machines on which they were developed, but may fail when executing without administrator privileges. For example, some programs will want to write configuration files (such as .INI files) in their installation directory under C:\Program Files, or to modify shared parts of the system registry. However, under most access control policies, that behavior is not allowed without administrator privileges.

Tip: In some cases, older software, or software that wants to write in a subdirectory of Program Files, can be made to work for unprivileged users without risk, if a root directory (outside Program Files) is specified as the installation directory for the application.
(Examples: I have found that this applies to Cincom Smalltalk 7.3.x, Allegro Common Lisp and Musicator 4.x)

International settings and character sets

International character sets: There are numerous schemes for representing non-English characters on computers, and this is a continuing challenge for international users. Some of us have simply resigned and avoid using non-English characters altogether, in file names and folder names. In recent years, the introduction of the Unicode 16-bit character set has led to a great improvement, but a large body of legacy software and legacy programmers (*sic*) is still in use. Most modern operating systems now support Unicode, as do mainstream office applications, but some scientific software still does not. Software exported from English-speaking countries often behaves erratically when users try to use national characters, especially in long filenames and pathnames.

Conversely, most European and Asian developers insist on presenting their user interfaces in English, even though some are not very good at it. Spelling errors and awkward grammar can make their products appear less than trustworthy to native speakers of English.

While this all has nothing to do specifically with lab software, it means that there may be a real incentive to prefer software from your own part of the world.

The decimal character. A fact that can have annoying consequences for your data processing is that different countries have different conventions for representing the decimal point (usually the period, but sometimes the comma). Professionally written software will query the operating system and use the decimal character from the locale information. Scientific software written in the U.S. is often oblivious to this issue, so non-US users may find that they have to convert their number tables between periods and commas. A workaround that sometimes works is to set the Windows Control Panel -> Regional settings -> Number formats to the US settings, but this will affect all your other applications as well.

In general, mainstream office programs are written to use the system-defined decimal character automatically. In recent versions only of Microsoft Excel, this issue has been recognized: you can override the default and specify another decimal character. While this is potentially even more confusing, it can be a help when you are exchanging spreadsheets with foreign sites, or are trying to use MS Excel with software into which the decimal character was hard-coded.

The list separator character. Following from the above, usually the comma is the list separator, but where the comma is employed as the decimal character, the semi-colon usually becomes the list separator.

To put this issue in perspective, the Microsoft handbook for software developers on how to localize Windows software for international markets runs to over eleven hundred pages. Few developers of lab software will read it all. [http://www.microsoft.com/mspress/books/5717.asp]

Hardware categories

The term "sound card" used to mean a circuit board carrying A/D and D/A converters and a small onboard analog mixer, to be inserted into the computer. Nowadays the term "sound card" seems to have been generalised to encompass all physical forms of audio converters, including

- audio circuits that are part of the motherboard chipset inside your computer,
- external units connected to the computer by USB or Firewire or Bluetooth or a LAN or a wireless LAN.

Sound cards that are potentially useful in research can be divided into three broad categories: consumer sound cards, pro audio cards, and data acquisition boards. All of these types are available as internal cards, external devices and even credit-card size PC-cards (formerly PCMCIA). A few high-end models come with a built-in digital signal processor (DSP) that can be custom-programmed by expert users.

Consumer sound cards

Consumer sound cards are built to be convenient to lay users as mass-market domestic appliances. They can have decent sound quality, and they can be very useful for some research tasks; but one must be aware of many potential problems, some of which are discussed below. Their audio specifications are usually very poorly documented or not documented at all.

Pro audio cards

Pro audio cards usually have excellent sound quality, and often have more than two channels of input and/or output, but still they universally lack the frequency response down 0 Hz (DC) that is necessary for non-audio signals such as those from physiological and aerodynamic transducers.

Data acquisition boards

Data acquisition boards do have DC response, and they solve most research problems, but they are built in smaller quantities and therefore are more expensive. They often require additional signal conditioning devices such as anti-aliasing filters. The same data acquisition boards that we used to mount in our PC/XT:s, before the advent of the multimedia PC, have now often been reissued as external devices with a USB connection. This is often a very good solution (see External hardware).

Data acquisition boards will usually tie you up to the software provided by the board manufacturer. Such software can be a bit clunky, so you need to make sure that the software does what you need in a workable way, and that the files that it saves are in a format that is easily read by other software. Data acquisition boards usually cannot be used with ordinary sound editing software, although there are exceptions. The Labview software system from National Instruments has gained a large market penetration and is supported by several hardware manufacturers. Data Translation, Inc. have a large range of boards with a unified software interface called DT-Open Layers.

Most data acquisition boards do come with a subroutine library, so that a software developer can control the board using any popular programming language. Therefore, the supplier will tell you that any particular function you need is easily accomplished with a bit of programming. If so, that is rarely a solution. Don't buy it, unless you know for sure that you can get the programming job done within your budget.

DSP systems

There are a few high-end hardware+software products that enable end users who are not DSP programmers to use hardware signal processors. While these are expensive, their specifications are detailed, and their performance is generally less susceptible to interference from other things that are going on in the host PC. These products enable you to build customized real-time signal-processing setups at a functional block diagram level. Examples include several modular systems from Tucker-Davis Technologies (US), and the Aladdin system from Hitech Development AB in Sweden. In addition to those, there are also systems that are targeted toward music creation, and possible to use in the lab (if you can live without DC response), such as SCOPE from Creamware, and KYMA/Capybara by Symbolic Sound Corporation.
There are also several DSP software-only systems that run 'natively' on the personal computer, without dedicated hardware: PureData, Max/MSP, Reaktor to name a few. These usually run only on sound cards that are supported by the operating system.

Basic requirements

In all cases, read the specs carefully, and make sure you get a board with

- 16-bit converters or better
- sufficient bandwidth (often called "throughput")
- bipolar inputs (not unipolar)
- adjustable or self-configuring anti-aliasing and reconstruction filters

If you will be acquiring physiological signals through galvanic contact to living bodies, the device needs to be certified for such use. Be sure to test the whole chain, from recording with the microphone and other transducers, all the way to a spreadsheet file of results.

External versus internal hardware

Internal sound cards (for desktop systems only)

Pros:

- high bandwidth, thanks to the parallel bus connection
- not likely to be borrowed
- no external controls that can be tampered with
- more stable temperature, if the computer is left on

Contras:

- require installation
- tied to one machine

Common problems with consumer sound cards

Below is a list of specific problems that can disqualify consumer-grade sound cards from research duty. In my opinion, this list is usually argument enough for buying a laboratory grade data acquisition board, rather than trying to work around the shortcomings of consumer sound cards.

Problems that usually do not apply to pro audio cards have been marked NPA (not pro audio) in this list.

Sound "enhancements"

Some sound cards come with extra features (in hardware or in software) that purport to enhance the sound, with bass boost, spatial surround emulation, reverberation, equalizers, and other distortions that have no place at all in the laboratory. Some may even have automatic gain control (AGC) when recording. Make sure that all such features, if present, are disabled.

Analog performance

The **analog section** of the sound cards can have various problems, including distortion/clipping, channel crosstalk, left/right channel reversal, and a residual DC offset. The DC offset may even change when the input gain control is changed. If you perform listening tests at high sound levels (with care), note that very few sound cards are able to drive directly a pair of headphones loudly without distortion (NPA)

The **signal-to-noise ratio** (SNR) can be poor. Fundamentally, the resolution (the bit width) of the A/D converters sets a theoretical limit on the SNR, which can be approximated in decibels as six times the number of bits on the digital side of the converter (e.g., 96 dB for 16-bit converters). However, most consumer cards fall far short of the theoretical performance and are far noisier (NPA).

The sound devices built into laptop computers can be particularly appalling in this regard. Sometimes the internal activities of the computer will cause interference in the sound, such as scratching noises when the mouse is moved or when a disk drive is working. When recording, there can be hum problems that become worse when the laptop is powered from the AC adapter. Reject such a machine.

How to tell: record silence, with the microphone turned off or disconnected, and inspect the result in a wave editor; and measure the level of the residual noise. Some residual noise will always be visible when the magnification is cranked up to a thousand times or more. If any noise is visible without magnification, the signal-to-noise ratio is woefully inadequate.

Analog overload. Because analog clipping does not sound quite so bad as driving the A/D-converter out of range, cheap consumer sound cards are sometimes designed intentionally to limit the output voltage swing of the input preamplifier on the card such that the A/D-converter that follows will never be driven out of range. This is very frustrating, because the analog clipping cannot be detected after the fact by the software.

How to tell: turn up the microphone preamp gain, record into a wave editor on the computer, and verify that you can make the signal clip in the editor, without any overload indicators blinking on the external analog gear. If you cannot, that is a problem. It is better to have the signal clip by exceeding the voltage range of the A/D converter, than to have it clip in the preceding analog electronics. It sounds worse, but it is more obvious that it is happening.

A good way to check for clipping of either kind is to make a full-range spectrogram of a recorded signal. Clipping will show up in black-on-white spectrograms as an distinct vertical dark band across all frequencies.

Low gain on the microphone input. Normal computer users usually use very close miking, as for dictation or Skype phoning or voice commands. Also, the high levels of electromagnetic noise in the vicinity of digital circuits make it difficult to keep the inherent noise down. Manufacturers therefore usually choose a microphone gain that is far too insensitive for research work. Always try to avoid using the microphone input of a consumer sound card. This is especially true on laptop computers. If a **line level input** is available, use it, with an external microphone preamplifier. However, in the year 2006 typically only a **mono microphone input** can be found on laptops. Best, use an external sound card with a (digital) USB or Firewire connection. USB microphones are also available, of varying quality.

No DC response. For physiological measurements, the analog inputs must have a frequency response down to 0 Hz. Audio cards **never** have this by default, only data acquisition boards do. A few audio cards can be custom-ordered with DC response, but this often makes the input unipolar (0…+x volts) rather than bipolar (-x…+x volts).

Voltage level mismatch to other instrumentation. Many laboratory instruments, such as preamplifiers for pressure or EGG transducers, will output voltages of up to +/- 10V or even more. This will overload the input on most sound cards, so attenuators are often needed. A simple 9:1 resistor ladder will attenuate the voltage by a factor of 10 (-20 dB), and will usually suffice. Data acquisition boards often have programmable gain, which helps, but it is important to use that facility properly.

Sampling rate inaccuracy. Consumer cards are not carefully checked in this regard, and few casual listeners will notice if the sampling rate is off by one or two percent. Such errors are rare, but we have known them; and for perceptual experiments e.g. on timing and pitch, such sampling rate errors can invalidate the results.
How to tell: Test the playback by downloading or creating audio files with tones of a known frequency and playing them to a laboratory grade frequency counter (e.g. 441 points per cycle played at 44100 Hz sampling rate should give a 100.00 Hz tone). The input and output converters usually run off the same internal digital clock. Therefore it is necessary, but not sufficient, to record a sine tone and check that it has the same frequency on playback (which perhaps takes place on different hardware), to the precision that you need for your work.

For historical reasons, the sampling rate 44100 Hz is sometimes implemented as 44097 Hz. This particular small difference is therefore not an error, and it is very rarely of any significance. It adds up to a time discrepancy of about one second over four hours of uninterrupted playback.

Audio peripherals

External sound cards

There are now a large number of audio interfaces that connect to the computer with USB or Firewire. This is usually a very good solution, if pro audio specifications are adequate. If you do field work, then try to get one with built-in microphone preamps, and that will take its power from the computer, through the USB lead, rather than through a separate mains adapter. Note that in some devices, even the maximum gain of the built-in microphone pre-amps can be on the low side.
Some interfaces have a digital S/P-DIF output, and more rarely also a S/P-DIF input. A digital input can be useful for re-archiving DAT tapes, as mentioned below.

Early USB 1.0 sound devices often had problems with glitching (clicks or interruptions), but with the advent of the much faster USB 2.0 protocol, this seems to have improved.

Microphones

Good sound starts with a good microphone. There are many things to consider when choosing and using a microphone, enough to warrant a separate paper. Jan Svec and Svante Granqvist are in the process of writing it.
[From Rahul Shrivastav:] Do not use microphones that come packaged with speech recognition software, as they often contain extra unspecified signal processing that serves to suppress background noise etc.

Loudspeakers

To hear details in voices, you need high-quality reproduction. The loudspeakers that come bundled with computers or are sold in computer warehouses are generally of very poor quality, or intended for a surround sound, or both. The loudspeakers that are sometimes built into flat panel displays are particularly awful. Discard them, and find a supplier of professional sound studio equipment. Buy a pair of small but high-quality self-powered monitor speakers. A curious rule of thumb is that if the loudspeaker enclosure is made of metal rather than plastic or board, the loudspeaker is probably good enough.

A desktop mixer

If you have multiple sound sources (computer, MiniDisc, CD player, etc), get a small desktop mixer with several stereo inputs. This is better than a hifi-amplifier with an input selector, because all the sources can be mixed in all the time.

If you want to record using such a desktop system, you need to watch out for howling feedback, which can occur in at least two ways:

1. Acoustic feedback: microphone - desktop mixer – loudspeaker - microphone
2. Electronic feedback: desktop mixer out - sound card mixer line in - sound card mixer line out - desktop mixer computer in

Standby power

Most audio components have AC power adapters that consume a standby power even when they are not turned on. To help the environment, arrange the AC power so that when you turn off the computer, the power strip for the ancillary components is switched off as well. There are special power strips with a built-in relay that powers up the outlets when a main device is turned on. Often they also have built-in surge protection.

Transferring DAT and MiniDisc recordings to computer storage

As I write this, digital audio tape (DAT) technology has passed its prime, but occasionally it will be desirable to retrieve and transfer material archived on DAT to optical disks, such as CD-ROM or DVD-ROM. MiniDisc recorders are still an interesting option for making field recordings, and the same need to transfer to computer exists for them. Ideally this transfer should be done digitally, without losing a single bit, but to do so requires a digital audio hardware interface on your computer. If you have a high-quality pro audio card, it may be adequate, and simpler, to do the transfer with an analog connection – but then level calibrations may be lost. The rest of this section is about achieving bit-correct digital transfer.

To keep us on our toes, the digital outputs on DAT and MiniDisc machines come in several different forms. There is a professional variant called AES/EBU, and at least two consumer variants, called S/P-DIF (Sony/Philips Digital Interface) and TosLink (from Toshiba). AES/EBU always uses balanced XLR connectors. S/P-DIF is usually unbalanced on RCA connectors, but other connectors are also used. TosLink is an optical fibre version of S/P-DIF. PUT PICTURES HERE. Converters that allow you to interconnect machines with different systems are available from pro audio suppliers.

On computers, consumer sound cards and laptops sometimes have S/P-DIF or TosLink outputs, but inputs are more rare. The knowledgeable consumer may want to play back digital audio from a PC over a hi-fi system, but the music industry did not want also to encourage error-free copying from the hi-fi to the PC. There are however several pro audio cards and also USB audio devices with coaxial and/or optical S/P-DIF inputs available from your local computer music store.

☠ A few consumer cards, notably some Soundblaster Live models from Creative Labs, have an input which is digital from the outside, but is always converted internally to analog for simple mixing with other sound sources, and then reconverted to digital, perhaps at a different sampling rate. This defeats the advantages of bit-correct digital copying.

Sampling rates: DAT can run at 48, 44.1 or 32 kHz (long play). DAT recorders default to 48 kHz rather than 44.1 kHz when recording. Audio CD:s are always sampled at 44.1 kHz, so if you have audio recordings that you want to save to audio CD:s, you may have to perform a → sampling rate conversion.

Selecting the sync source. Normally your sound card clocks its converters using its own onboard crystal oscillator. However, digital audio over S/P-DIF requires that the receiver be synchronized to the source. In this case, your sound card is the receiver and the DAT player is the source. Some pro-audio cards with S/P-DIF inputs require you explicitly to set the S/P-DIF input as their clock source before you record from this input. This is typically done with some software control panel that is installed together with the card's driver software. If you don't, only silence will be recorded.

Compatibility: To pack a lot of information onto a small tape, the DAT technology was specified to very tight tolerances. Unfortunately not all DAT recorders meet these tolerances, so you may find that a DAT-format tape recorded on one machine will not replay reliably on another, or may even refuse to play at all. Data loss during DAT replay typically sounds like dropouts combined with a buzz. If you encounter this, the tape machine may need maintenance, or, try playing the tape on a different machine.

Side information: The AES/EBU data protocol allows for supplementary information to be inserted into the data stream. Some machines, especially in film and broadcasting, use this feature to store so-called SMPTE time code and time-of-day information on the tape itself, which can be very useful in the clinic. This information is usually not retrievable on consumer machines.

There are also multi-channel **digital instrumentation recorders**, for example, the TEAC PCM-200. Such machines allow the acquisition of many channels with a frequency response down to DC; they replaced the FM tape machines of the 1980's. The tapes look identical to audio DAT tapes, but their formatting is completely different, and specific to each manufacturer. Such tapes will not play back on standard audio DAT machines. To transfer such information bit-correctly, you must use the original model of instrumentation recorder, and also the hardware computer interface recommended by the manufacturer, often a so-called GPIB industry standard instrumentation interface.

Using sounds in Microsoft Powerpoint

There are several ways of playing sound examples as part of a Powerpoint presentation. In all cases, **test the audio** before the starting the presentation itself.

Although you don't see it, Powerpoint effectively invokes the Windows Media Player to play the sound formats it knows about, including MIDI files and MP3 files. This is convenient, but you might need to understand what is going on. The term "sound clip" is sometimes used to denote a piece of audio-only material, regardless of its format.

Sound clips can be embedded into the PPT file itself. While this is convenient when the presentation will be moved to another computer, it can make the PPT file very large. A sound clip can also be inserted as a link to a separate file. This file would normally be in the same folder as the PPT file, but it can reside anywhere: on another disk drive (or memory card, or CD/DVD player, etc) or even on the Internet. This can be effective if you often reuse the same examples in different presentations, but if you move to another computer, or if the medium is removed, the presentation is likely to break. In any case, you probably need to be aware of whether your examples are in fact embedded or not. If you use the Pack-and-Go feature of Powerpoint, even externally linked files may be copied and packed into the resulting standalone presentation file.

☠ CAUTION

The fidelity of your sound examples can be important, such as when you want to demonstrate a subtle timbre difference between two sounds. When preparing a presentation with audio examples, you must be aware that Powerpoint will sometimes **automatically compress embedded audio** quite brutally, without even telling you. It is not uncommon, on playback of embedded sound clips, to notice a lack of high frequencies as well as aliasing distortion that was not there in the sound file that you thought you inserted. This will be a major concern if you plan to use Powerpoint to run a listening test.

Tip: Prevent embedding

Powerpoint will normally try to embed small sound clip files, but not big ones. You can prevent embedding altogether, by setting a zero threshold for the maximum size of sound files to be embedded. (Tools -> Options -> General (tab) -> Link sounds with file sizes greater than <u>0</u> kB).

Tip: Make an audio CD of sound examples

One way of definitely avoiding compression is to store your sound examples as separate tracks on a separate audio CD, and to use the Powerpoint Action feature that plays a given CD track. To do this, you must remember to bring the CD itself, the presentation PC must have a CD drive available, and the "CD" input must be active in the Playback Settings of the Volume Control applet. This may be a good option if you have numerous long sound examples that take up a lot of space. The CD drive may take several seconds to rev up and find the track and play it, so you need to allow for this potentially annoying delay in your presentation. Or, of course, you can play the CD manually on a separate sound system – but then the sound is no longer connected to Powerpoint.

Tip: Use a separate media player program to control audio playback

Sometimes you may want to browse a sequence of slides while a list of sound examples is playing, without interrupting the sounds. Or, you may want to seek to a certain point in a sound file, and perhaps repeat it. These things can be done using a media player program (the Windows one, or a third-party player). Make a special playlist for your media player, containing the sound examples in the right order. Then set the option to display the media player as "always on top" of other windows, and have its window display in the most compact mode. Start the media player before Powerpoint, or, define a Powerpoint button to start the media player, with the appropriate playlist, at a convenient point in the presentation. Now the media player will float over your slides until you close it, and you can access both programs at the same time.

Oh, and check that the "random shuffle" option of the media player program is turned **off**. That one really had me baffled, playing my sound examples in random order…

Tip: Insert a shortcut to the Volume Control applet

Sometimes sound examples need their volume adjusted during a presentation, or the left-right balance needs to be changed to compare two simultaneous sounds. By including an action button that runs SNDVOL32.EXE (the Volume Control), you can get the mixer panel onto the screen and quickly access these controls, without switching out of your presentation. Remember, though, that the generic Volume Control program does not have an "always on top" option, and it will hide behind the slide show when you click onward. You can use ALT+TAB to get it back in front.

Tip: define shortcut keys to run other programs

Some laptops and desktop keyboards have user-definable keys that can run any given program when pressed. This can be very useful for bringing up other programs, including media players or the Volume Control. Furthermore, in Windows, *any* program can be assigned a CTRL+ALT+LETTER shortcut: find the EXE file, right-click on it and choose Properties, and find the tab with Shortcut in it.

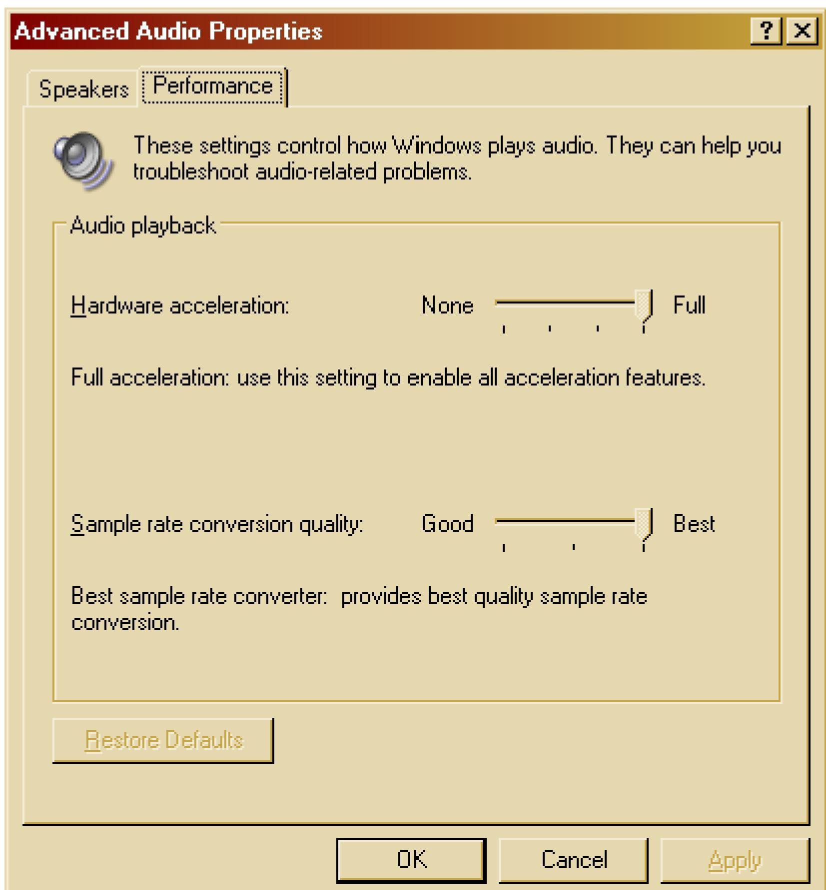
The operating system

Windows and its K-mixer

Most researchers will not need to delve into the details of the audio subsystem architecture, but there is one issue which is very important to recognize. From Windows 98 and onwards, Windows includes a software kernel component kalled the K-mixer, which is all but inaccessible to users. The purpose of the K-mixer is to cater to the situation of several multitasking programs that simultaneously want to play audio to the user. Different programs may want to play sounds at different sampling rates, different numbers of channels, and different bit depths. For example, your e-mail program might signal new mail with a "plong" while you are doing your French lessons and maybe listening to a webcast radio channel. The user does not want to know, but wants the sounds to be heard even if they have to be mixed. The K-mixer therefore accepts all requests to play sounds, converts them on the fly to some format that it deems appropriate, and mixes them together onto the system's audio output device.

While this "enhances the user experience," it is problematic for the researcher and also for the broadcaster, because it means that you cannot really know that your signals have not been tampered with along the way. For example, you might think that you are acquiring DAT digitally at 48 kHz, but your sound card will happily play it back at 44.1 kHz – after K-mixer has converted the sampling rate. Worse, the conversion can occur during recording, and so it was not bit-correct, after all.

The only control that Windows offers the user is a curious slider to control the amount of CPU effort that is to be spent on the sampling rate conversion, see figure. We are not told what it does. It has the euphemistically labelled positions Good-Better-Best. That is all. And the Good setting is not actually good, as you can imagine. I have had poor conversion happen when it was set to Best.



Using legacy speech software on Windows

One example of when the k-mixer might be a problem is when you work with a specialist or legacy signal processing system, such as ILS or CSL, but you listen to the files (perhaps converted to .WAV format) using a sound card under Windows. The special systems often run at non-consumer sampling rates such as 10, 12, 16, 20, 25 or 50 kHz; and therefore Windows will probably resample the output to 44.1 kHz or 48 kHz. If the spectrum is critical, particularly at high audio frequencies, check the spectrum of the sound *after* the D/A-converter, by re-recording and analyzing it. If it is not what you expect (e.g., if aliased partials are present), you may want either (1) to get a better sound card; or (2) determine the preferred native sampling rate of your sound card, and resample the files yourself to that rate before playback, using a reliable software utility.

Third-party audio drivers

The only really controlled workaround for the Windows k-mixer is the one that was developed by a frustrated computer music industry. It is to use a set of independent third-party audio drivers. Examples are so-called ASIO drivers, for which the protocol has been laid down by Steinberg, Inc.; or DirectKS. By using such drivers, you bypass most or all of the audio architecture in Windows. Both the applications and the hardware must be specifically compliant. Sounds originating in the Windows system will not be routed to that hardware.

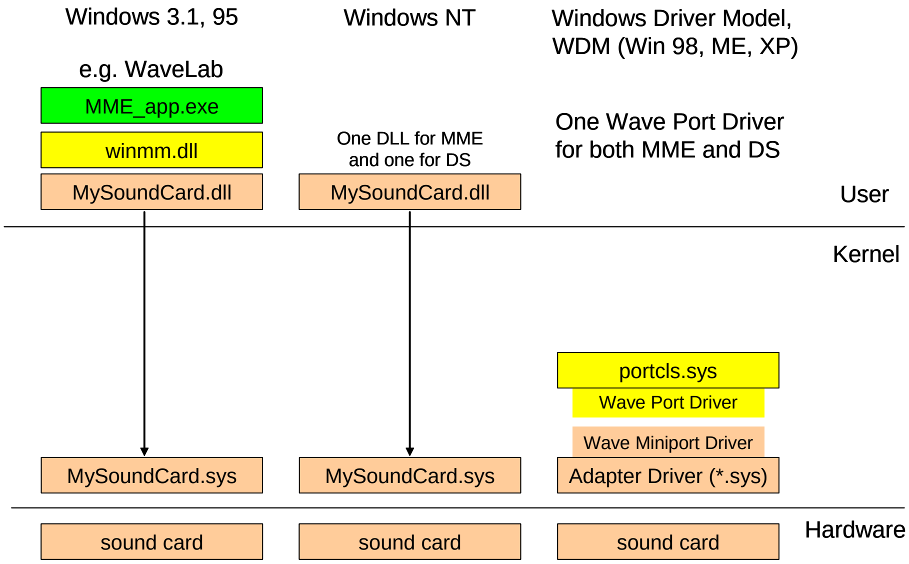
For the nerds

The audio subsystem has changed significantly with almost every major release of Windows. Several acronyms jostle for attention: MME, DirectX, WDM, UAA. The figures below, which were compiled by Jonas Ekeroot of the Swedish Broadcasting Corporation, suggest the complexity of the software architecture.

Windows Vista comes with a new major rehaul of the audio support. An interesting change is that the native format for digital sound will change from 16-bit integer to 32-bit floating point. This will have little consequence for the consumer, but it could mean an improvement for research applications. It remains to be seen to what extent the need for backward compatibility will add to the confusion.

To my knowledge, audio support in Windows Mobile (Pocket PC) is much reduced, and barely emulates the MME drivers only.

Below, "User" and "Kernel" refer to the two privilege levels at which the CPU may be executing in any given microsecond. Code that executes in User mode is prevented from accessing critical system functions and other programs' memory spaces. In Kernel mode, "anything goes", so coding mistakes can crash the whole system; but it runs faster.

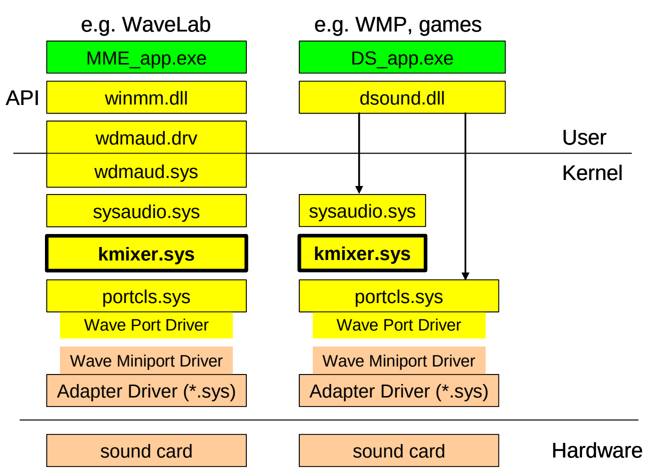


Older architectures compared to WDM

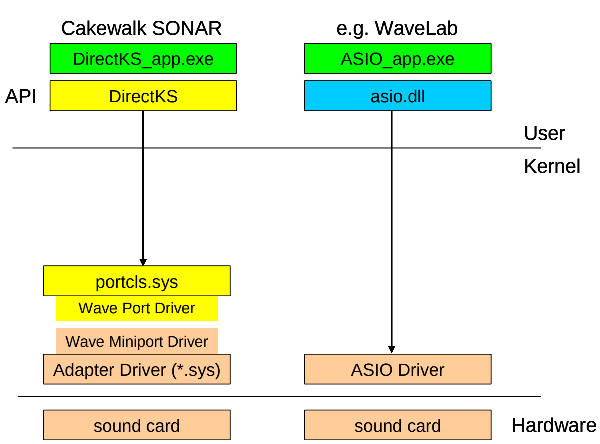
MME =MultiMedia Extensions, from the early 1990's, still in use.

DS = DirectSound, part of DirectX

Control is transferred from one module to another through subroutine calls.

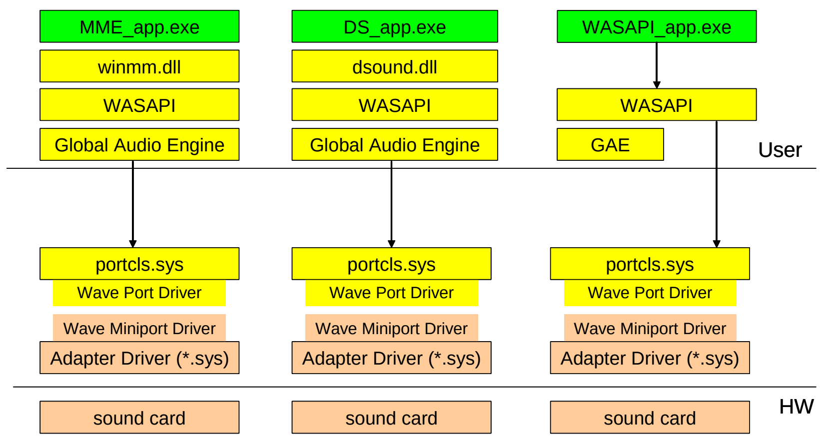


Current audio architecture under the Windows Driver Model, WDM, for MME and DirectSound applications.



Third-party workarounds that avoid the k-mixer: DirectKS and ASIO. Used with specialist applications such as for audio production work. ASIO drivers are hardware specific; and are often available also for Mac and Linux.

Windows Vista: WASAPI (Windows Audio Session API)



In Windows Vista, a new session API is introduced. The mixing functions have moved from Kernel to User mode.

Linux

Personally I have no experience of this, but there is much useful information that comes with the free audio-oriented Agnula distribution of Linux, at www.demudi.org . It also includes a wealth of free audio- and music-related software.

Macintosh OS X

Here I draw a blank, and welcome input from others.

Links

http://www.smcnetwork.org/

http://www.tascam.com/Products/US-428/W2k_XP_Optimize.pdf This document from 2002 is becoming a bit dated, but still contains a lot of good advice on how to optimize your PC for audio.

http://www.jakeludington.com/ask_jake/20050225_optimize_your_pc_for_audio_and_video.html

http://www.speech.kth.se/software/ Free downloads from KTH, including WaveSurfer and RTsect.

Correspondence

From Dr Eric Hunter at NCVS Denver.

I had a thought for the paper of recording pitfalls. It is on the line of archiving recordings. Two things we always deal with: (1) space/size of recordings, and (2) quality checking of recording and archives.

What we use is a program called FLAC, http://flac.sourceforge.net/

It is a lossless audio (wav) compression. It also has a check-sum that allows to check an archive. So we will 'flac' a wave file, and then burn it to CD (usually 2 cd's, duplicate everything...). Then we can check the file on the CD to see if the burn went ok. We have noticed that occasionally even when the cd writing software indicates that the file write went fine, there is a problem with the file that was written. The 'flac' check-sum catches this for us.

Anyway, maybe this will be useful in the document you write.

-Eric

Eric Hunter, Ph.D.
Assistant Research Scientist
National Center for Voice and Speech
The Denver Center for the Performing Arts
1101 13th St. | Denver, CO 80204
ehunter@dcpa.org | eric.hunter@ncvs2.org
Phone: 303-446-4839 | Fax: 303-893-6487

Voice analysis software packages - freeware

Wavesurfer (built on Snack, which is built on Tcl/Tk) www.speech.kth.se
Praat

Voice analysis software packages – supported products

Wevosys: www.wevosys.com LingWAVES
Kay Pentax: www.kaypentax.com CSL
Laryngograph: www.laryngograph.com
Saven Hitech: www.savenhitech.se Soundswell Core, Soundswell Voice

General signal analysis/processing products

Tucker-Davis Technologies (USA) www.tdt.com
OROS (France)
Aladdin Interactive DSP (Sweden) www.hitech.se/development
KYMA/Capybara from Symbolic Sound Corporation (USA) (mostly for music)
Labview (USA)

Matlab, with Signal Processing Toolbox and other add-ons (USA).
Note that a freeware, low-latency multichannel audio interface for Matlab will soon be available at KTH.

PureData www.puredata.org (freeware)

Max/MSP