Hang the DJ: Automatic Sequencing and Seamless Mixing of Dance-Music Tracks

Dave Cliff
Publishing Systems and Systems Laboratory
HP Laboratories Bristol
HPL-2000-104
9th August, 2000*

E-mail: dave_cliff@hp.com

user interfaces, music information retrieval, music sequencing, mixing, Disk-Jockey, automation, dance music, compilation

Many radio stations and night-clubs employ Disk-Jockeys (DJs) to provide a continuous stream or "mix" of music, built from a sequence of individual song-tracks. In the last decade, commercial pre-recorded compilation CDs of DJ mixes have become a booming market. DJs exercise skill in deciding an appropriate sequence of tracks and in mixing 'seamlessly' from one track to the next. Online access to large-scale archives of digitized music via automated music information retrieval systems offers users the possibility of discovering many new songs they like, but the majority of consumers are unlikely to want to learn the DJ skills of sequencing and mixing. This paper describes an automatic method by which compilations of dance-music can be sequenced and seamlessly mixed by computer, with minimal user involvement. The user may specify a selection of tracks, and may give a qualitative indication of the type of mix required. The resultant mix can be presented as a continuous single digital audio file, whether for burning to CD or for play-out from a virtual jukebox or personalized virtual radio station.

* Internal Accession Date Only Approved for External Publication

Hang the DJ: Automatic sequencing and seamless mixing of dance-music tracks

Dave Cliff
Digital Media Systems Department
Hewlett-Packard Laboratories
Bristol BS34 8QZ
England
dave_cliff@hp.com
Phone +44 117 312 8189

June 2000

Abstract

Many radio stations and night-clubs employ Disk-Jockeys (DJs) to provide a continuous stream or "mix" of music, built from a sequence of individual song-tracks. In the last decade, commercial pre-recorded compilation CDs of DJ mixes have become a booming market. DJs exercise skill in deciding an appropriate sequence of tracks and in mixing 'seamlessly' from one track to the next. Online access to large-scale archives of digitized music via automated music information retrieval systems offers users the possibility of discovering many new songs they like, but the majority of consumers are unlikely to want to learn the DJ skills of sequencing and mixing. This paper describes an automatic method by which compilations of dance-music can be sequenced and seamlessly mixed by computer, with minimal user involvement. The user may specify a selection of tracks, and may give a qualitative indication of the type of mix required. The resultant mix can be presented as a continuous single digital audio file, whether for burning to CD or for play-out from a virtual jukebox or personalized virtual radio station.

Topic Area:
User interfaces for music information retrieval.

Keywords:
Automation; Music Sequencing; Music Mixing; Disk-Jockey; Dance Music; Compilation.

1. Introduction

What will happen when the major problems in music information retrieval are solved? Imagine if they were solved now, so the 250,000-plus songs held by Mp3.com could be automatically ranked in order of similarity to your entire record collection, or maybe your current favorite five songs. The resultant ranking would be a personalized music recommendation service based not on the purchasing patterns of strangers, but on your personal taste in music. This could be a good way of finding new music to listen to.

Say such a recommendation service came up with a bunch of songs. How would you want them presented to you? Maybe streamed over the web as a "virtual radio" channel, or maybe burnt onto a CD. However, most young(ish) people listening to radio, or dancing to CDs, want their songs to have been 'mixed' by a disk-jockey (DJ). The job of a DJ isn't simply just playing a bunch of records. There's art and skill in deciding the order of the records, and in mixing between successive records.

For these reasons, many radio stations and night-clubs employ DJs to provide a continuous stream or "mix" of music, built from a sequence of individual song-tracks. Moreover, sales of commercial compilation CDs of DJ mixes (a type of CD unknown until 1992: Brewster & Broughton 1999, p.368) are currently booming, constituting a major sector of chart CD sales (in the UK at least). The London *Ministry of Sound* nightclub is estimated to have income from sales of its compilation CDs (produced by its own *Sound of Ministry* independent record label) in excess of £20m ($30m) per year (Kershaw 2000, p.60). The shelf-life of a typical compilation CD is short (often no more than 6 months), but in that time it may sell 500,000 copies (Kershaw 2000, p.60).

In recent years, DJ's have become a new breed of music performer. Top DJs are international stars, earning millions of dollars. The fee a top DJ receives for producing a compilation CD (a task that may take little more than a couple of hours) may be up to £50,000 ($75,000). Kids who want to be cool want to be DJs: sales of DJ equipment now exceed sales of guitars in the UK. Nevertheless, working as a DJ requires skill at two levels: the macro-level of *sequencing* and the micro-level of *mixing*.

Sequencing involves deciding an appropriate ordering of tracks. While this is manifestly dependent on the DJ's personal taste in music, there is an element to sequencing that is somewhat more mechanistic. In many instances, the music's tempo (traditionally measured in units of "bpm": beats-per-minute) will be systematically and smoothly varied over the duration of the DJ's playing session (which typically lasts anything from 40 minutes to 6 hours). The tempo is dynamically varied to follow some trajectory, in a manner analogous to the distinct movements that constitute a symphony in classical music. In a night-club, there will be definite periods of "warm-up" (when the tempo of the tracks rises over time – encouraging the clientele onto the dance-floor), plateaus (keeping the dancers dancing) and peaks (aimed at driving the dancers into a brief frenzy, after which they need to buy another drink). Toward the end of a DJ session, there may be a period where the tempo is progressively reduced (the "come-down" or "chill"), to start to encourage people to think about leaving, or buying another drink. Commercial DJ-mixed compilation CDs almost always follow some such trajectory – sometimes split over multiple disks.

The micro-level of mixing 'seamlessly' from one track to the next depends on artful "cross-fading": fading down the volume of the outgoing track while simultaneously bringing up the volume of the incoming track. At some point during a cross-fade, both tracks will be audible simultaneously: this works best if the two tracks are playing at the same tempo and with no phase difference: so-called "beat-matching", which allows one track to be faded into the next without any discernable alteration in the underlying rhythmic beat. Figure 1 shows the effects on the output mix of a poorly executed cross-fade with no beat-matching, while Figure 2 shows the results of a well-executed cross-fade.

For this reason, seamless mixing also often requires dynamic alteration in the pitch, tempo, and phase of the two tracks being mixed between. Alterations in pitch and tempo are achieved by reducing or increasing the playback speed of a track, while phase differences are rectified by very briefly pausing playback of one of the tracks. Sometimes it is not possible to beat-match two tracks because even when their tempo is identical and there is no phase difference, their interaction sounds bad. In such cases the DJ may choose to cross-fade at a point where the beat is absent in one of the tracks – that is, during a so-called "breakdown" in the beat of the track.

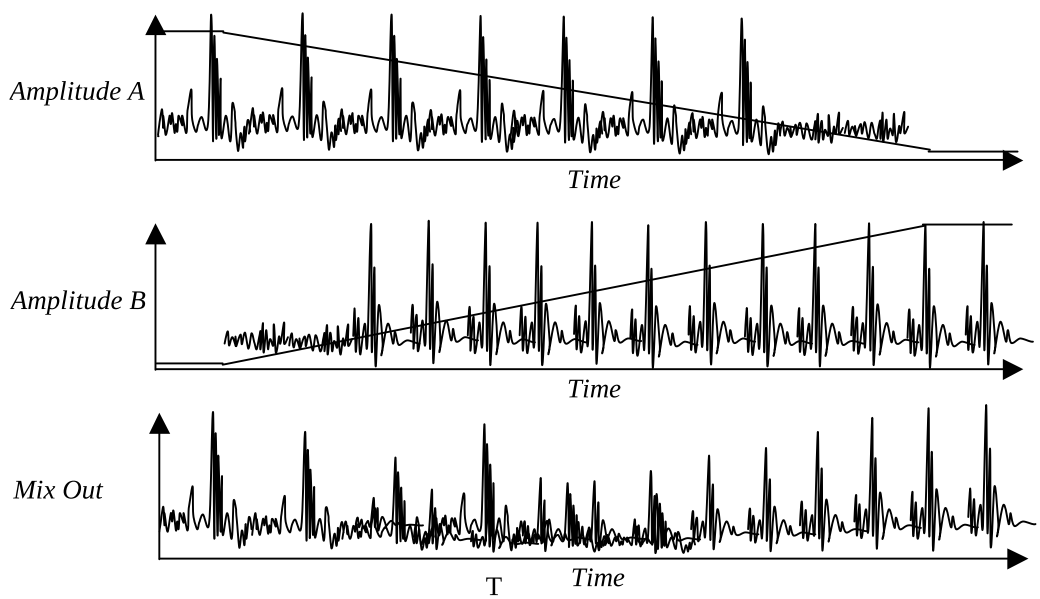


***Figure 1. Cross-fading done badly**. The upper two graphs show amplitude-time plots of the audio in two songs being cross-faded:* A *is the outgoing track and* B *(with a faster tempo) is the incoming track. If the amplitude is signal strength following low-pass filtering, then the pronounced peaks are likely to represent the songs' underlying beat. The dashed diagonal lines show the relative volumes of tracks* A *and* B*: note that the sum of the two volumes is constant during the cross-fade. Note also that the beats in* A *and* B *are only coincident at time* T *(indicated by the vertical dotted line). The bottom graph shows the resulting mixed output. Because the beats in* A *and* B *are not coincident elsewhere, there is a noticeable drop in the amplitude of the beats in the mix. Also, around time* T *the beats in the two tracks combine to give a brief section in the output mix where there is an audible beat-pattern that is aperiodic and that has approximately twice the tempo of* A *and* B.

This paper reports on a process that automates the DJ's task, and which could be used as a component of a user interface in commercial music information retrieval systems. It describes a method for specifying a collection of tracks that are then automatically sequenced to follow some tempo trajectory, and then seamlessly mixed without any need for further human intervention and without any need for human preprocessing of the tracks.

In particular, the method described here operates on "dance" styles of music, where the rhythmic element of the music is regular and pronounced. In the sublime poetry of English Law, these styles are defined in Clause 58 (1) (b) of the 1994 Criminal Justice Act as "…sounds wholly or predominantly characterized by the emission of a succession of repetitive beats." Such music styles include those popularly known as "disco", "electronica", "house", "garage", "techno", "hip-hop", "drum n bass", and "trance". These styles are the mainstay of many nightclubs and of dance-oriented radio stations, and they regularly constitute the majority of the songs in the national top-twenty charts of many countries.

Relevant prior work is described in Section 2. Section 3 then describes the new method in detail.

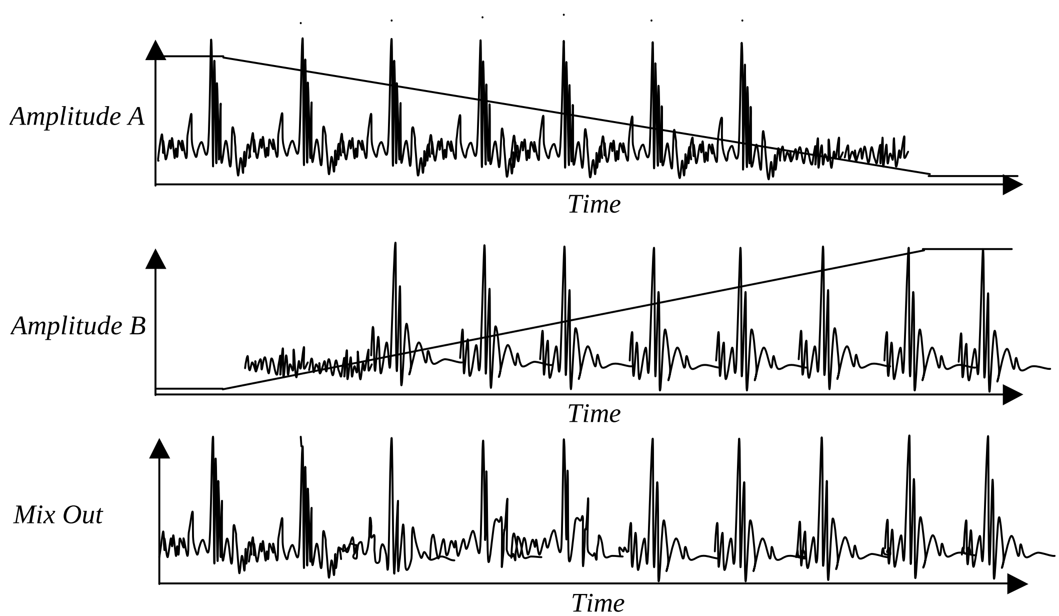


***Figure 2. Cross-fading done well**. As in Figure 1, the upper two graphs show amplitude-time plots of the audio in two songs being cross-faded where* A *is the outgoing track and* B *is the incoming track. But here track* B *is time-stretched (e.g. by slowing its playback speed) so that its tempo matches that of track* A, *and the two tracks are synchronized such that there is no phase difference between their beat patterns. The two tracks beat-match for 6 beats (indicated by the vertical dotted lines – note that in practice the beat-matching would last for many more beats). In consequence, the mix output shows no discernable drop in amplitude, and shows a constant beat tempo as the two tracks are cross-faded. Once the cross-fade is complete, the playback speed of track* B *may be gradually increased (reducing the time-stretch back to zero).*

2. Related Work

The European patent application entitled *Automatically performed crossover between two consecutively played back sets of audio data* (L'Hopital, 1999) claims the invention of a solution to the DJ problem, but has the following disadvantages:

- It requires pre-specified "begin" (end-of-fade-in) and "end" (start-of-fade-out) cue-markers to be added to each track's audio data. It gives no indication of any automatic method for doing this, and so the only reasonable interpretation is that skilled human operators are employed to decide on these begin and end points for each and every track.

- Each track has only one "begin" and one "end" marker, whereas in most situations the end of fade-in and the start of fade-out for any one track will depend on the circumstances of its usage (i.e., the particular sequence it is being used in, and its location within that sequence).

- In the third claim of L'Hopital's patent, varying the speed of playback over the "begin" or "end" periods of a track is claimed as an aspect of the invention. Yet no method or apparatus is specified or claimed for dealing with the nontrivial effects that variations in playback speed routinely have on the pitch, tempo, and phase of the tracks being mixed between.

- It says nothing about ordering of tracks within an extended sequence of tracks (i.e., more than two) and the temporal evolution (trajectory) of music tempo that skilled DJ's devise in such extended sequences.

A commercial product called *Databeat DJ Master* is marketed by Sound Management Services Ltd of Newbury, UK, to bars and pubs (see http://www.databeat.com). At the time of writing, Databeat is installed in over 1000 sites around the world, with remote updating of each installation from the Databeat archive. All music in the Databeat system is catalogued by human operatives who record production data (such as year of release) along with data used by their proprietary mixing software. This mixing data includes the start-chord, end-chord, track tempo (bpm), and the location of (human-placed) "begin" and "end" cue-points similar to those involved in the method claimed by L'Hopital. Thus, the Databeat system is not fully automatic in that it requires human operatives to generate the cataloging meta-data. Details of how the human-generated meta-data is employed by the Databeat automatic DJ software are not available.

3. Hands-free automatic Djing.

Starting with access to a collection of songs stored as digital audio files (in any format – mp3, wav, etc), the operation of a version of our method can be summarized as follows. It takes as input a list of desired tracks (which may have been specified by the user, or may come from another source such as an automatic recommendation service, or a random picker). This list of *n* tracks is referred to here as the *set*.

The first steps in the process involve determining a sequence for the set, where the degree of user involvement in the sequencing process is variable from fully user-specified to fully automatic. The digital audio tracks do not require any pre-processing to locate fade begin and end points, because these points are calculated dynamically for each sequence and indeed the fade-in and fade-out points for any one track are likely to vary from sequence to sequence. We use prior art digital signal processing (DSP) techniques to automatically vary the pitch and tempo (i.e., the playback speed) of tracks as appropriate to the particular sequence, and the process then automatically sets the relative phase of successive tracks with high precision, to ensure seamless beat-matched mixing. The resultant continuous large file of digital audio can be produced as output for subsequent recording (e.g. burning onto CD) or playout (e.g. over audio broadcast or narrowcast systems). Additional data, such as the time-points at which one track transitions to another, may also be recorded (e.g. so as to provide a table of contents for a CD to be written with time indices for each track). Individual steps in the process are described below. Further details are available elsewhere (Cliff, 2000). The process is described here as a linear sequence of steps, but in Section 4 we discuss nonlinear versions.

3.1 Track Mapping

Beat-detection techniques known in prior art (e.g. Yamada et al, 1997) are used to determine a *tempo-map* for each of the tracks to go into the mix. The tempo-map is an indication of the bpm measured at intervals across the duration of the track. Figure 3 shows a schematic illustration of the beats in a sample of music and the corresponding tempo-map.

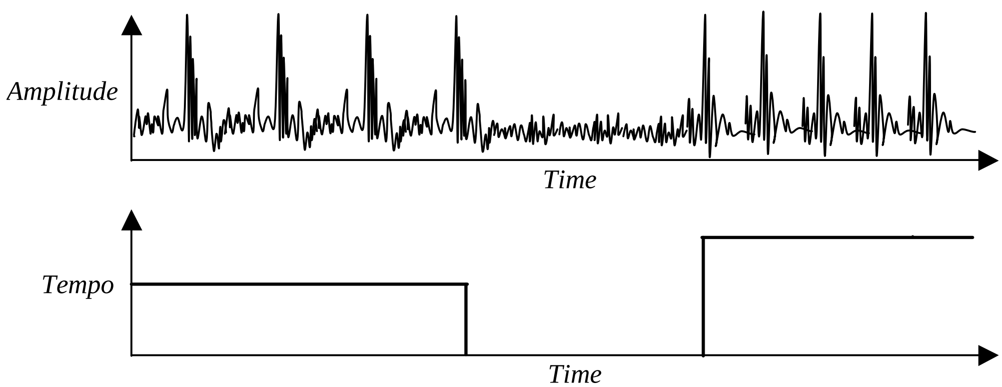


Figure 3:tempo-map. *The upper graph shows a schematic amplitude-time plot for a section of a song where a tempo change occurs following a "breakdown". The lower graph schematically illustrates a corresponding tempo-map showing the initial lower tempo, followed by the breakdown (zero tempo), followed by the subsequent return of the beat at a higher tempo.*

Similarly well-established DSP techniques can be used to determine maps of amplitude and possibly also pitch/key for each track. These maps are dependent only on the original recorded version of the track, and so could be saved for the next time the track is used, or could all be computed in advance for each track in the music collection.

3.2 Trajectory Specification

The sequence of tracks can be fully and explicitly specified by the user, or sequencing can be completely automated, or it can be partially automated with some guidance from the user. This guidance can take the form of the user specifying a qualitative tempo trajectory (QTT) and optionally also by specifying some ordering constraints (e.g. “don’t play Track A before Track B”). A QTT is a specification of how the tempo should vary over the duration of the mix, expressed in relative, rather than absolute, terms. This allows the same QTT to be used when compiling separate mixes of different durations, or of different tempo-ranges. For instance, a simple “warm-up” QTT would show a monotonic increase in tempo from a minimum value at the start of the mix to a maximum value at the end of the mix. A graphic representation of this would be to plot a straight upward-sloping line on a graph of tempo over time: example QTTs are illustrated in Figure 4. Significantly, the duration of the mix is not explicitly specified, so the same QTT could be used for a mix lasting 30 minutes, and for one lasting 3 hours. Similarly, the bpm values of the minimum and maximum tempos in the mix are also unstated, thereby allowing the same QTT to be used for mixing both a compilation where all tracks have tempos in the range 100-120bpm, and for one where the set's tempo range is 125-145bpm.

The QTT for a mix might be directly specified by the user, or chosen by the user from a set of pre-specified QTTs, or randomly chosen by the system from that set of pre-specified QTTs. The user may also specify a maximum time duration for the mix (e.g., in preparing a mix to be burnt to a standard-format CD, the duration should be no longer than 74 minutes).

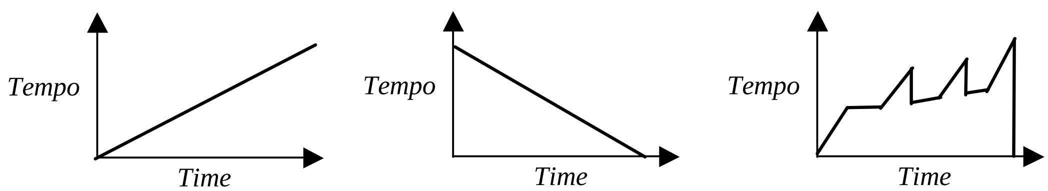


Figure 4: Qualitative Tempo Trajectories (QTTs). *The left-hand graph shows a QTT for a "warm-up" set. The center graph shows a QTT for a "come-down" set. The right-hand graph shows a QTT suitable for a protracted set on radio or in a nightclub: after the initial warm up comes a plateau that is followed by a sequence of three peaks of successively higher maximal tempo, with the set ending immediately after the fastest song.*

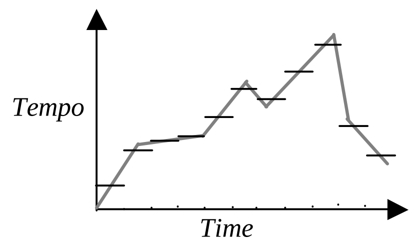
3.3 Sequencing

The QTT imposes constraints on the sequence of the tracks, constituting a partial ordering. For example, the (qualitative) point in the mix where the lowest tempo is specified on the QTT indicates the approximate location of the slowest track in the set; and the point where the highest tempo appears in the QTT indicates the approximate location of the fastest track in the set.

Turning these approximate indications into a concrete sequence is a straightforward procedure. The QTT is discretized by dividing it into *n* sections. And the tempos of these QTT sections are ranked in order from highest to lowest. The tracks in the set are also sorted in order of their overall tempo, from highest to lowest (a track’s overall tempo is taken as the average of the nonzero tempos recorded over track’s tempo-map). These two ordered lists are then used to determine the sequence, with the highest-tempo track being assigned to the highest-tempo QTT section, the second-highest-tempo track being assigned to the second-highest tempo QTT section, and so on, as illustrated in Figure 5.

Elementary constraint-satisfaction techniques can be used to check for violations of any of the user-supplied ordering constraints and to take appropriate action when violations are detected.

The end result is a list of the tracks in the set, in the order they are to appear in the mix: this list is the sequence for the mix.



Track	*Tempo*
B	140
G	138
L	138
E	137
H	136
K	136
F	135
C	131
A	130
D	128
J	127

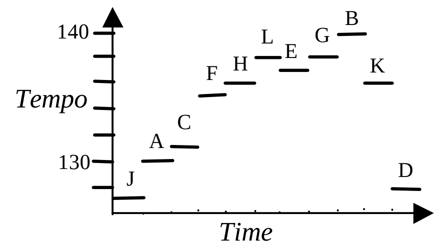


Figure 5: Sequencing. *Left: the QTT is discretized by dividing it into* n *slots (*n=*11 here). Center: the* n *tracks are ranked by tempo. Right: the highest-tempo-ranked track is assigned to the highest-tempo QTT slot; the second-highest-tempo-ranked track assigned to the second-highest-tempo QTT slot, and so on until the lowest-tempo-ranked track is assigned to the lowest-tempo QTT slot. The final sequence of tracks in this example is thus* J-A-C-F-H-L-E-G-B-K-D.

3.4 Overlapping.

In order for the mix to be "seamless", there should be no "dead-spots" between tracks. While the avoidance of absolute silences is trivial, it is insufficient because many dance-music tracks have long (and relatively boring) "intro" (start) and "outro" (end) sections, where often the main melody or vocal content is absent, with only the rhythmic component of the song being present. Few listeners would want to hear the outro of one track playing to its very end, followed by the intro of the next track played from its very beginning. Indeed, the intention of the music producers is that these intro/outro sections are to be played while cross-fading from/to the outgoing/incoming track in the mix. Thus, the tracks in the mix have to be overlapped.

Determining the degree of overlap between tracks depends on whether the user has specified a maximum duration for the mix. If no duration has been specified, an initial arrangement of overlaps can be set by making each track overlap with the next by some pre-specified amount – either a fixed number of seconds, or a number of seconds that is a fixed proportion of that track's duration. If a mix-duration of d seconds has been specified, and the total combined length of the n tracks in the mix is l seconds, then the initial arrangement of overlaps can be set by overlapping each pair of tracks by $(l-d)/(n-1)$ seconds. Note that this assumes that $d<l$ (if $d>=l$ then the duration set by the user is irrelevant, and the overlap is set as if no duration was specified).

Once the initial overlaps have been determined, fine-tuning can be performed by examining the tempo-maps for each pair of tracks in the areas where they overlap. If an initial overlap occurs near to a position where either track shows a beat "breakdown", the overlap point may be moved to allow the cross-fade to occur during the breakdown. If a maximum duration has been specified for the mix, moves that lengthen the mix-duration are forbidden. Once any such moves have been specified, the tempo-maps for the tracks are combined to create an overall tempo-map for the entire mix.

3.5 Time-stretching and Beat-matching.

Combining the tempo maps reveals areas in the mix where overlapping tracks have different tempos. The tracks in these areas are then time-stretched so that their tempos of the tracks in the overlapped portion are approximately the same. Equality of tempo is desirable but often unachievable because it may be necessary to time-stretch in such a way that a track's tempo alters as the track progresses. For example, in a three-track set where the track tempos are 100, 110, and 120 bpm, the middle track could be stretched so that its tempo is 100bpm when it is cross-faded in over the first track and then increases slowly to 120bpm by the time it is cross-faded out under the final track.

Once the time-stretching has been applied to bring all the tempos into line, simple beat-detection algorithms can be applied to identify the positions of the beats in the tracks and to align overlapping tracks such that there is no (or minimal) phase difference between them. This involves moving the tracks in the sequence by small amounts of time – typically less than half a second. Finally the volumes of the tracks are altered in the overlap areas, in a manner analogous to the cross-fading volume alterations a DJ performs. In the simplest case, linear amplitude decay/increase modulates the outgoing/incoming track, but other curves for these amplitude envelopes are possible.

4. Nonlinearities

Although the process as described above is linear, starting with a list of tracks to go in the set and progressing through the stages described in Sections 3.1 to 3.5, there are obvious ways in which the process could be altered to be nonlinear or iterative. A nonlinear process could be invoked if the user specifies only a small number of songs relative to the time limit on the mix and also requests that the unspecified time is filled with tracks chosen by the system. In this case, it may be more appropriate to introduce "wild-card" (unassigned) tracks when sequencing and overlapping, and to then select appropriate songs from some song database to instantiate these wild-cards after sequencing is complete. Deferring automatic selection of songs in this way allows the system's choice of songs to be constrained by the tempo and/or pitch of the surrounding tracks. In particular, the deferred instantiation of wild-card tracks can be used to bridge over major tempo transitions in the sequence. For example, if the user's specifications and choices result in a 140bpm song having to be mixed into a 100bpm song with beat-matching during the cross-fade, both tracks would require unacceptably high alterations in their playback speeds. (Slowing the fast track by setting its playback speed to 86% reduces its tempo to 120bpm and speeding the slow track by an extra 20% increases its tempo to 120bpm also, but the songs are likely to sound unappealingly different from the familiar original versions at these playback speeds). In such cases it may be better to add a small number of wildcard tracks between the two user-specified tracks: these could be chosen from a song database on the basis of their tempo. In the example just given, if three wildcard tracks are introduced, constrained to have tempos of around 130, 120, and 110bpm, then the tempo changes between successive tracks in the final mix would all require less extreme (and hence more tolerable) alterations in playback speed.

5. Conclusion

The process described here goes some way towards replacing the task performed by human DJs. It has potential use as a component in the user-interface to consumer music information retrieval systems, converting the output of such systems from a set of tracks into a continuous seamless mix. Such mixes are suitable for playout over streaming media (e.g., in personalized internet radio) or for writing to an appropriate recording medium (such as CD or flash ROM card) for subsequent playback.

References

Brewster, B., & F. Broughton. *Last Night a DJ Saved My Life*. London: Hodder Headline, 1999.

Cliff, D. *Automatic Compilation of Songs*. European patent application 00303960.9, May 2000.

Kershaw, M. Compilation Wars. *Mixmag* 104, pp.58—64, January 2000.

L'Hopital, C. *Automatically performed crossover between two consecutively played back sets of audio data.* European patent application EP0932157A, July 1999.

Yamada, Y., T. Kimura, T. Funada, & G. Inoshita. *Apparatus for detecting the number of beats.* US Patent 5,614,687, March 1997.