



UNIVERSIDADE CATÓLICA PORTUGUESA

**AUTOMATIC MIXING SYSTEMS USING ADAPTIVE DIGITAL
AUDIO EFFECTS**

Tese apresentada à Universidade Católica Portuguesa
para obtenção do grau de Doutor em Ciência e Tecnologia das Artes — Informática Musical

por

Pedro Duarte Leal Gomes Pestana

ESCOLA DAS ARTES
Fevereiro 2013



UNIVERSIDADE CATÓLICA PORTUGUESA

**AUTOMATIC MIXING SYSTEMS USING ADAPTIVE DIGITAL
AUDIO EFFECTS**

Tese apresentada à Universidade Católica Portuguesa
para obtenção do grau de Doutor em Ciência e Tecnologia das Artes — Informática Musical

Por Pedro Duarte Leal Gomes Pestana

Sob orientação do Professor Doutor Álvaro Barbosa
e Dr. Joshua D. Reiss

ESCOLA DAS ARTES
Fevereiro 2013

Abstract

The field of Automatic Mixing has been ripe with research over the past five years. There are now systems that emulate level and panning decisions with some sophistication, and these have been evaluated as being close in quality to offerings produced by a professional sound engineer. However, a real-world mix is an integrated effort between several axes, and we are missing a ground-truth on how expert mixers holistically perform their art and craft.

The purpose of this thesis is to fill that knowledge gap in what mixing best practices are. An exhaustive number of approaches were made to complete this endeavor, starting with a review of academic, technical and non-technical literature. A significant number of interviews with expert professional mixers followed, which led to a quantitative questionnaire in best practices. Simultaneously, over 20 different subjective evaluation tests were performed with the collaboration of medium-skilled to expert listeners. Another validation approach was through the devising of algorithms that could extract significant content from full-mixes. The pursuit of these goals involved the creation of two extensive datasets, one of multi-track unmixed material, and a second one of commercially very successful songs to be analyzed for production patterns.

All these concurrent efforts led to the creation of a shifting collection of assumptions that crystalized into its final form as the 88-assumption database that is the kernel of this thesis. Unlike previous efforts, it focuses on top-quality, offline, studio mixing of music. For each individual assumption we have undertaken a validation process whenever possible, and summarized efforts done so far in terms of implementation. As a part of the testing and validation process, many implementation algorithms were developed, some of them novel, and are described together with the assumptions that inspired them.

The work was organized so that it may contribute as a self-contained blueprint to many of the upcoming approaches to computer-assisted mixing. Despite the considerable length of our effort, we acknowledge that there is so much more to unravel on this topic, that the present work serves only as a modest foundation stone, probably opening up more questions for further research than the ones it now closes.

Resumo

O campo de investigação em técnicas de mistura automática de conteúdo musical tem sido foco de considerável actividade ao longo dos últimos cinco anos. Embora conceptualmente o tópico tenha já originado implementações diversas desde a década de 70 do século passado, é apenas com o acesso a velocidades rápidas de processamento computacional que pode ser exportado para o universo dos efeitos de audio digitais, especificamente dentro da categoria de efeitos adaptativos cruzados.

A mistura automática tem necessariamente de ir buscar inspiração à sua congénere manual, pois tem sido esta a criadora de todos os paradigmas de apresentação sonora pós-composição e pós-arranjo ao longo do último século. No entanto, o conjunto de boas-práticas em termos de técnicas de mistura tradicionais não está devidamente documentado.

Este trabalho procura ser um passo inicial para colmatar devidamente essa lacuna, tentando responder à questão “*qual é o conjunto de regras-base seguidas por engenheiros de som de nomeada, e que possam ser adaptadas a sistemas automáticos?*” A criação de uma verdade de fundo é o primeiro de muitos passos na direcção de um sistema que possa ser considerado inteligente, mas tem sido negligenciada até agora, e substituída por um conjunto de especulações e investigações pontuais, muitas vezes erróneas.

A introdução subreptícia do termo ‘técnicas’ não é inocente, visto que a mistura é largamente considerada uma tarefa parcialmente artística ou criativa. Não é nosso objectivo imiscuirmo-nos nesse domínio, e restringiremos a nossa discussão à parte técnica ou correctiva. A nossa abordagem torna-se também diferente do que tem sido estudado dentro do tópico de mistura automática pois é focada na mistura de música em estúdio (a maior parte da pesquisa tem sido sobre sistemas de mistura ao vivo) e como tal aceitamos estratégias que não sejam exequíveis em tempo real. Criámos também o conceito acessório de *mistura assistida por computador*, em que (em algumas áreas) não pretendemos imitar o trabalho de um técnico de som, mas sim estender as suas possibilidades através de processos tão minuciosos que nunca poderiam ser executados por um sistema não automático.

É bastante frequente considerar uma divisão funcional da mistura audio em eixos dimensionais, um esquema que guia a nossa abordagem quantitativa. Imaginamos assim um eixo ligado à amplitude ou volume apercebido, um eixo horizontal de posição panorâmica, um eixo vertical de distribuição espectral e um eixo de profundidade que inclui processamento temporal e outras pistas discretas de distância e posição. A estes eixos que podem ser visualizados num espaço imaginado, acresce a dimensão temporal, em conceitos de micro-dinâmica e macro-dinâmica.

Partimos de um enquadramento agnóstico, em direcção à criação de um conjunto de pressupostos ou hipóteses sobre quais são as regras consensuais em cada eixo. Como a literatura existente é exígua e de natureza bastante especulativa, o nosso processo inicial para listagem de potenciais premissas foi levado a cabo através de um conjunto extenso de entrevistas em formato aberto. Criámos uma colecção de critérios (explicados no Apêndice A) sobre o que determinava a caracterização de um técnico de mistura em topo de carreira, e utilizámos uma estratégia de selecção baseada em amostragem aleatória para criar um painel de especialistas, cuja opinião poderia contribuir para o desenvolvimento do processo normativo (estes

especialistas estão também listados no Apêndice A). Esta etapa, aumentada por algumas reflexões informadas pelo desenvolvimento paralelo de modelos que executámos, gerou uma colecção de 88 pressupostos a investigar, que formam o núcleo deste trabalho.

De acordo com uma ordenação de importância, investigámos a fundo cada um dos assuntos, incluindo revisão bibliográfica, análise de exemplos reais, e seguimento no processo de entrevistas. À medida que a envolvente teórica cristalizava tornou-se urgente criar um conjunto de medidas de validação que nos permitissem avaliar o grau de interesse ou universalidade de cada um dos pressupostos.

Foram estabelecidos cinco processos de validação, não mutuamente exclusivos:

1. Medição objectiva de parâmetros de misturas de sucesso comercial em formato finalizado.
2. Análise de resultados de testes subjectivos efectuados por um grupo alargado de utilizadores com uma experiência moderada.
3. Criação de um questionário quantitativo a preencher pela lista de engenheiros de som reputados, e passível de um tratamento estatístico.
4. Análise de um conjunto de exercícios de mistura completados por engenheiros de som reputados.
5. Medição subjectiva de parâmetros de misturas de sucesso comercial em formato finalizado.

O conteúdo de cada pressuposto dita o tipo de processo de validação mais adequado a utilizar, e em alguns casos foram utilizados vários em simultâneo. O processo mais poderoso é claramente a medição objectiva de parâmetros de misturas, mas como um projecto finalizado apenas permite acesso a um ficheiro estereofónico, os parâmetros analisáveis são reduzidos, e esta foi uma estratégia que utilizámos apenas em 9 pressupostos. Os testes subjectivos, utilizados numa amostra suficientemente expressiva, oferecem resultados por vezes claros e quantitativos. Executámos mais de 20 testes diferentes que ofereceram respostas para 36 pressupostos. O problema desta estratégia é principalmente o custo de execução, mas entendemos também que alguns aspectos são demasiado ténues para serem avaliados correctamente, e outros são irreproduzíveis em formato teste, por serem multi-dimensionais. O passo seguinte (não do ponto de vista temporal, mas do ponto de vista de importância de resultados) foi a elaboração do questionário que foi respondido por 49 especialistas mundiais, e esclareceu 56 pressupostos. Embora seja um instrumento de valor irrepreensível em alguns tópicos, e o número de respostas nos dê bastante confiança nos resultados quando estes se apresentam claros, esta estratégia de validação começa a apresentar problemas de viés mais complexos: o frasear da pergunta pode ser determinante; os especialistas que não respondem podem ter uma prática radicalmente diferente; os especialistas que responder podem estar inclinados a omitir informação que é importante para o seu sucesso, e assim por diante. Tomámos os maiores cuidados para evitar todos estes problemas, mas o facto é que neste ponto a validação perde alguma da sua objectividade. Propusemos também um exercício detalhado de mistura de duas músicas a 5 dos profissionais que colaboraram nas entrevistas e questionários. Os resultados são fascinantes e poderosos, mas o tamanho amostral é claramente insuficiente, visto que este é um teste quase impossível de executar dada a sua exigência em termos de tempo. O exercício serviu para clarificar 18 pressupostos, embora nunca o tenhamos apresentado como um resultado taxativo ou estatisticamente significativo. Finalmente, 2 pressupostos foram sujeitos a medições subjectivas sobre misturas comerciais. Há que lembrar que o processo

de entrevistas e a revisão bibliográfica é também em si uma estratégia de validação, embora claramente mais subjectiva.

Um outro tópico importante do trabalho foi a descrição de sugestões de implementação para os pressupostos nos quais tenham emergido. Fizemos uma revisão do que já tinha sido implementado em sistemas automáticos, e chegámos à conclusão que 28 dos pressupostos já tinham sido utilizados ou mal utilizados em projectos anteriores. Em algumas das áreas fizemos contribuições relevantes do ponto de vista de implementação (por exemplo na área de reverberação automática, onde simplesmente não existe nada implementado), e noutras as estratégias de implementação serviram o processo de validação, auxiliando a construção de testes subjectivos. Ao longo da elaboração da tese criámos dois *datasets* musicais: um de conteúdos multi-pistas até agora inexistente, e que serviu os testes subjectivos e alguns testes informais; e outro com *singles* que tenham sido número um no Reino Unido ou Estados Unidos entre 1950 e 2010, e que serviu as medições objectivas de parâmetros sobre misturas finalizadas. O primeiro encontra-se disponível *online* para pesquisa académica.

O trabalho foi organizado de forma a poder contribuir como uma planta arquitectónica auto-contida para a construção das abordagens que se seguirão dentro do tema de mistura automática e mistura assistida por computador. Pese embora seja um trabalho de uma dimensão ambiciosa, reconhecemos que ainda há tanto a conhecer sobre o assunto, que o presente esforço constitui apenas uma modesta pedra de fundação, e provavelmente abre mais questões para pesquisa futura do que aquelas que ficam agora fechadas.