



República Federativa do Brasil  
Ministério do Desenvolvimento, Indústria  
e do Comércio Exterior  
Instituto Nacional da Propriedade Industrial.

(21) **PI0619197-5 A2**



\* B R P I O 6 1 9 1 9 7 A 2 \*

(22) Data de Depósito: 27/11/2006  
(43) Data da Publicação: 20/09/2011  
(RPI 2124)

(51) *Int.Cl.:*  
H04H 60/37  
H04H 60/58  
H04N 7/16

(54) **Título:** DETECÇÃO DE CONTEÚDO DE REPETIÇÃO EM MÍDIA DE DIFUSÃO

(30) **Prioridade Unionista:** 29/11/2005 US 60/740,760, 29/08/2006 US 60/823,881, 29/08/2006 US 60/823,881, 29/11/2005 US 60/740,760

(73) **Titular(es):** Google Inc.

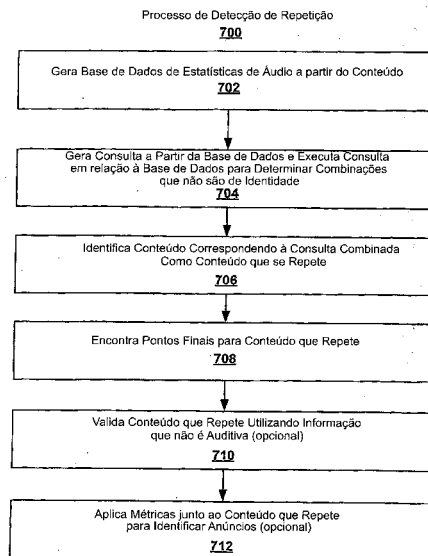
(72) **Inventor(es):** Michael Fink, Michele Covell, Shumeet Baluja

(74) **Procurador(es):** Dannemann, Siemsen, Bigler & Ipanema Moreira

(86) **Pedido Internacional:** PCT US2006045549 de 27/11/2006

(87) **Publicação Internacional:** WO 2007/064640 de 07/06/2007

(57) **Resumo:** DETECÇÃO DE CONTEÚDO DE REPETIÇÃO EM MÍDIA DE DIFUSÃO. A presente invenção refere-se aos sistemas, métodos, dispositivos e produtos de programa de computador que provêem aplicações sociais e interativas para a detecção de conteúdo de repetição em uma mídia de difusão. Em algumas implementações, um método inclui: a geração de um banco de dados de estatísticas de áudio a partir de um conteúdo; a geração de uma consulta a partir do banco de dados de estatísticas de áudio; a rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade; se uma combinação não de identidade existir, a identificação de conteúdo correspondente à consulta combinada como um conteúdo de repetição.



Relatório Descritivo da Patente de Invenção para "**DETECÇÃO DE CONTEÚDO DE REPETIÇÃO EM MÍDIA DE DIFUSÃO**".

PEDIDOS RELACIONADOS

5 Este pedido reivindica o benefício de prioridade a partir do Pedido de Patente Provisória U.S. Nº 60/740.760, para "Environment-Based Referrals", depositado em 29 de novembro de 2006, cujo pedido é incorporado como referência aqui em sua totalidade.

10 Este pedido reivindica o benefício da prioridade do Pedido de Patente Provisória U.S. Nº 60/823.881, para "Audio Identification Based on Signatures", depositado em 29 de agosto de 2006, cujo pedido é incorporado como referência aqui em sua totalidade.

15 Este pedido está relacionado ao Pedido de Patente U.S. Nº \_\_\_\_\_, para "Determining Popularity Ratings Using Social and Interactive Applications For Mass Media", depositado em 27 de novembro de 2006, Protocolo Legal Nº GP-672-00-US/16113-0630001, e ao Pedido de Patente U.S. Nº \_\_\_\_\_, para "Social and Interactive Applications For Mass Media", depositado em 27 de novembro de 2006, Protocolo Legal Nº GP-636-00-US/16113-060001. Cada um destes pedidos é incorporado aqui como referência em sua totalidade.

20 CAMPO TÉCNICO

As implementações descritas estão relacionadas a aplicações sociais e interativas para mídia de massa.

ANTECEDENTES

25 Os sistemas de televisão convencional e de televisão interativa carecem da capacidade de detecção de redifusões de propaganda embutida em uma programação de televisão. Os dispositivos de gravação convencionais permitem que os usuários armazenem programas de televisão (incluindo comerciais) para redifusão em uma data ou horário posterior. Uma reclamação comum dentre as difusoras é sua incapacidade de lucrar por estas redifusões, o que da perspectiva das difusoras contabiliza uma propaganda  
30 "grátis" para os anunciantes que compraram espaço na transmissão original do espetáculo.

## SUMÁRIO

As deficiências descritas acima são consideradas pelos sistemas, métodos, aparelhos, interfaces de usuário e produtos de programa de computador descritos que detectam um conteúdo de repetição em uma mídia de difusão.

Em algumas implementações, um método inclui: a geração de uma consulta a partir de um banco de dados de estatísticas de áudio; a rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade; e se a combinação não de identidade existir, a identificação do conteúdo correspondente à consulta combinada como um conteúdo de repetição.

Em algumas implementações, um sistema inclui um processador e um meio que pode ser lido em computador operativamente acoplado ao processador. O meio que pode ser lido em computador inclui instruções, as quais, quando executadas pelo processador, fazem com que o processador realize as operações de: geração de uma consulta a partir de um banco de dados de estatísticas de áudio; a rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade, onde as estatísticas de áudio são geradas a partir de um conteúdo; e se a combinação não de identidade for encontrada, a identificação do conteúdo correspondente à consulta combinada como um conteúdo de repetição.

Outras implementações são dirigidas a sistemas, métodos, aparelhos, interfaces de usuário e produtos de programa de computador.

## 25 DESCRIÇÃO DOS DESENHOS

A Figura 1 é um diagrama de blocos de uma modalidade de um sistema de personalização de massa.

A Figura 2 ilustra uma modalidade de um sistema de identificação de áudio ambiente, incluindo uma interface de lado de cliente descrita na Figura 1.

A Figura 3 é um fluxograma de uma modalidade de um processo para a provisão de aplicações de personalização de massa.

A Figura 4 é um fluxograma de uma modalidade de um processo de impressão digital de áudio.

A Figura 5 é um fluxograma de uma modalidade de interface de usuário para interação com aplicações de personalização de massa.

5 A Figura 6 é um diagrama de blocos de uma modalidade de arquitetura de hardware para um sistema de cliente implementando a interface de lado de cliente descrita na Figura 1.

A Figura 7 é um fluxograma de uma modalidade de um processo de detecção de repetição.

## 10 DESCRIÇÃO DETALHADA

### Aplicações de Personalização de Massa

As aplicações de personalização de massa provêem uma informação personalizada e interativa relacionada a difusões de mídia de massa (por exemplo, televisão, rádio, filmes, difusões pela Internet, etc.). Essas aplicações incluem, mas não estão limitadas a: camadas de informação personalizada, comunidades de par social *ad hoc*, classificações de popularidade em tempo real e favoritos de vídeo (ou áudio), etc. Embora alguns dos exemplos de mídia de massa descritos aqui sejam no contexto de difusões de televisão, as implementações descritas são igualmente aplicáveis a difusões de rádio e/ou música.

As camadas de informação personalizada provêem uma informação complementar ao canal de mídia de massa. Os exemplos de camadas de informação personalizada incluem, mas não estão limitados a: moda, política, negócios, saúde, viagens, etc. Por exemplo, enquanto assiste a um segmento de notícias sobre uma celebridade, uma camada de moda é apresentada ao espectador em uma tela de televisão ou em um dispositivo de exibição de computador, o qual provê uma informação e/ou imagens relacionadas às roupas ou acessórios que a celebridade está usando no segmento de notícias. Adicionalmente, as camadas personalizadas podem incluir anúncios promovendo produtos ou serviços relacionados ao segmento de notícias, tal como um link para uma loja de roupas que esteja vendendo as roupas que a celebridade está usando.

As comunidades de par social *ad hoc* provêm um local para comentário entre usuários que estejam assistindo ao mesmo show na televisão ou ouvindo a mesma estação de rádio. Por exemplo, a um usuário que esteja assistindo às últimas manchetes da CNN pode ser provido um meio de comentário (por exemplo, uma sala de bate-papo, um quadro de mensagem, uma página da wiki, um link de vídeo, etc.), que permita que o usuário bata papo, comente ou leia as respostas de outros espectadores para a difusão de mídia de massa em andamento.

As classificações de popularidade em tempo real provêm aos provedores de conteúdo e usuários uma informação de classificação (similar às classificações de Nielsen). Por exemplo, a um usuário podem ser providas instantaneamente classificações de popularidade em tempo real de canais de televisão ou estações de rádio sendo assistidas ou ouvidas pela rede social do usuário e/ou por pessoas com demografia similar.

Os favoritos de vídeo ou áudio provêm aos usuários formas de pouco esforço de criação de bibliotecas personalizadas de seu conteúdo de difusão favorito. Por exemplo, um usuário pode simplesmente pressionar um botão em um computador ou em um dispositivo de controle remoto e um fragmento de áudio ambiente e/ou vídeo do conteúdo de difusão é gravado, processado e salvo. O fragmento pode ser usado como um favorito para referência ao programa ou a porções do programa, para visualização mais tarde. O favorito pode ser compartilhado com amigos ou salvo para referência pessoal futura.

#### Rede de Personalização de Massa

A Figura 1 é um diagrama de blocos de um sistema de personalização de massa 100 para a provisão de aplicativos de personalização de massa. O sistema 100 inclui uma ou mais interfaces de lado de cliente 102, um servidor de banco de dados de áudio 104 e um servidor de aplicativo social 106, todos os quais se comunicando por uma rede 108 (por exemplo, pela Internet, uma intranet, uma LAN, uma rede sem fio, etc.).

Uma interface de cliente 102 pode ser qualquer dispositivo que permita que um usuário introduza e receba uma informação, a qual é capaz

de apresentar uma interface de usuário em um dispositivo de exibição, incluindo, mas não limitando: um computador de mesa ou portátil; um dispositivo eletrônico; um telefone; um telefone móvel; um sistema de exibição; uma televisão; um monitor de computador; um sistema de navegação; um toca-  
5 dor/gravador de mídia portátil; um assistente digital pessoal (PDA); um console de jogos; um dispositivo eletrônico portátil; e um dispositivo ou aparato eletrônico embutido. A interface de cliente 102 é descrita em maiores detalhes com respeito à Figura 2.

Em algumas implementações, a interface de cliente 102 inclui  
10 um detector de áudio ambiente (por exemplo, um microfone) para a monitoração e a gravação do áudio ambiente de uma difusão de mídia de massa em um ambiente de difusão (por exemplo, uma sala de estar de um usuário). Um ou mais segmentos de áudio ambiente ou "fragmentos" são convertidos em sumários estatísticos distintivos e robustos, referidos como "impressões  
15 digitais de áudio" ou "descritores". Em algumas implementações, os descritores são arquivos comprimidos contendo um ou mais componentes de assinatura de áudio que podem ser comparados com um banco de dados de descritores de referência gerados previamente ou estatísticas associadas à difusão de mídia de massa.

20 Uma técnica de geração de impressões digitais de áudio para identificação de música é descrita em Ke, Y., Hoiem, D., Sukthankar, R. (2005), *Computer Vision for Music Identification*, *In Proc. Computer Vision and Pattern Recognition*, o qual é incorporado aqui como referência em sua totalidade. Em algumas implementações, a abordagem de identificação de  
25 música proposta por (referido a partir deste ponto como "Ke et al.") é adaptada para a geração de descritores para dados de áudio de televisão e consultas, conforme descrito com respeito à Figura 4.

Uma técnica para a geração de descritores de áudio usando ondulações é descrita no Pedido de Patente Provisória U.S. Nº 60/823.881,  
30 para "Audio Identification Based on Signatures". Este pedido descreve uma técnica que usa uma combinação de técnicas de visão em computador e algoritmos de processamento de transmissão contínua de dados em larga

escala para a criação de descritores compactos/impressões digitais de fragmentos de áudio que podem ser eficientemente combinados. A técnica usa ondulações, o que é uma ferramenta matemática conhecida para decomposição de forma hierárquica de funções.

5                    Em "Audio Identification Based on Signatures", uma implementação de um processo de recuperação inclui as etapas a seguir: 1) dados os espectros de áudio de um fragmento de áudio, extrair imagens espectrais, por exemplo, de  $11,6 \times w$  de duração, com um espaçamento randômico em média de  $d$  ms. Para cada imagem espectral: 2) computar ondulações na  
10 imagem espectral; 3) extrair as  $t$  ondulações de topo; 4) criar uma representação binária das  $t$  ondulações de topo; 5) usar um mini-hash para a criação de uma subimpressão digital das  $t$  ondulações de topo; 6) usar LSH para  $b$  intervalos e  $l$  tabelas de hash para encontrar segmentos de subimpressão digital que sejam combinações próximas; 7) descartar as subimpressões  
15 digitais com menos de  $v$  combinações; 8) computar uma distância de Hamming a partir das subimpressões digitais candidatas remanescentes para a subimpressão digital de consulta; e 9) usar uma programação dinâmica para as combinações combinadas ao longo do tempo.

Em algumas implementações, os descritores e um identificador  
20 de usuário associado ("ID de usuário") para a identificação da interface de lado de cliente 102 são enviados para o servidor de banco de dados de áudio 104 através da rede 108. O servidor de banco de dados de áudio 104 compara o descritor com uma pluralidade de descritores de referência, os quais foram previamente determinados e armazenados em um banco de  
25 dados de áudio 110 acoplado ao servidor de banco de dados de áudio 104. Em algumas implementações, o servidor de banco de dados de áudio 104 continuamente atualiza os descritores de referência armazenados no banco de dados de áudio 110 a partir de difusões de mídia de massa recentes.

O servidor de banco de dados de áudio 104 determina as melho-  
30 res combinações entre os descritores recebidos e os descritores de referência e envia a informação de melhor combinação para o servidor de aplicativo social 106. O processo de combinação é descrito mais plenamente com res-

peito à Figura 4.

Em algumas implementações, o servidor de aplicativo social 106 aceita conexões de navegador da web associadas à interface de lado de cliente 102. Usando a informação de melhor combinação, o servidor de aplicativo social 106 agrega a informação personalizada para o usuário e envia a informação personalizada para a interface de lado de cliente 102. A informação personalizada pode incluir, mas não está limitada a: anúncios, camadas de informação personalizada, classificações de popularidade e uma informação associada a um meio de comentário (por exemplo, comunidades de par social *ad hoc*, fóruns, grupos de discussão, videoconferências, etc.).

Em algumas implementações, a informação personalizada pode ser usada para a criação de uma sala de bate-papo para espectadores, sem se saber qual show os espectadores estão assistindo em tempo real. As salas de bate-papo podem ser criadas diretamente pela comparação de descritores nas transmissões contínuas de dados transmitidas pelos sistemas clientes para a determinação de combinações. Isto é, salas de bate-papo podem ser criadas em torno de espectadores tendo descritores combinando. Em uma implementação como essa, não há necessidade de comparação dos descritores recebidos a partir dos espectadores em relação aos descritores de referência.

Em algumas implementações, o servidor de aplicativo social 106 apresenta uma página da web para a interface de lado de cliente 102, a qual é recebida e exibida por um navegador da web (por exemplo, o Microsoft Internet Explorer™) rodando na interface de lado de cliente 102. O servidor de aplicativo social 106 também recebe o id de usuário a partir da interface de lado de cliente 102 e/ou do servidor de banco de dados de áudio 104, para ajudar na agregação de um conteúdo personalizado e apresentação de páginas da web para a interface de lado de cliente 102.

Deve ser evidente que outras implementações do sistema 100 são possíveis. Por exemplo, o sistema 100 pode incluir múltiplos bancos de dados de áudio 110, servidores de banco de dados de áudio 104 ou servidores de aplicativo social 106. Alternativamente, o servidor de banco de dados

de áudio 104 e o servidor de aplicativo social 106 podem ser um servidor ou sistema único, ou fazer parte de um recurso e/ou serviço de rede. Também, a rede 108 pode incluir múltiplas redes e links operativamente acoplados em conjunto em várias topologias e arranjos, usando-se uma variedade de dispositivos de rede (por exemplo, centros de conexão, roteadores, etc.) e meios (por exemplo, cobre, fibra ótica, frequências de rádio, etc.). As arquiteturas de cliente – servidor são descritas aqui apenas como um exemplo. Outras arquiteturas de computador são possíveis.

### Sistema de Identificação de Áudio Ambiente

10                   A Figura 2 ilustra um sistema de identificação de áudio ambiente 200, incluindo uma interface de lado de cliente 102, conforme descrito na Figura 1. O sistema 200 inclui um sistema de mídia de massa 202 (por exemplo, um aparelho de televisão, rádio, computador, dispositivo eletrônico, telefone móvel, console de jogos, aparato de rede, etc.), um detector de áudio ambiente 204, uma interface de lado de cliente 102 (por exemplo, um computador de mesa ou laptop, etc.) e um dispositivo de acesso de rede 206. Em algumas implementações, a interface de lado de cliente 102 inclui um dispositivo de exibição 210 para apresentação de uma interface de usuário (UI) para se permitir que um usuário interaja com um aplicativo de personalização de massa, conforme descrito com respeito à Figura 5.

20                   Em operação, o sistema de mídia de massa 202 gera áudio ambiente de uma difusão de mídia de massa (por exemplo, um áudio de televisão), o qual é detectado pelo detector de áudio ambiente 204. O detector de áudio ambiente 204 pode ser qualquer dispositivo que possa detectar áudio ambiente, incluindo um microfone independente e um microfone que seja integrado com a interface de lado de cliente 102. O áudio ambiente detectado é codificado pela interface de lado de cliente 102 para a provisão de descritores identificando o áudio ambiente. Os descritores são transmitidos para o servidor de banco de dados de áudio 104 por meio do dispositivo de acesso de rede 206 e da rede 108.

30                   Em algumas implementações, um software de cliente rodando na interface de lado de cliente 102 monitora continuamente e grava arquivos

de áudio de  $n$  segundos (por exemplo, de 5 segundos) ("fragmentos") de áudio ambiente. Os fragmentos então são convertidos em  $m$  quadros (por exemplo, 415 quadros) de descritores codificados de  $k$  bits (por exemplo, 32 bits), de acordo com um processo descrito com respeito à Figura 4. Em algumas implementações, a monitoração e a gravação são baseadas em evento. Por exemplo, a monitoração e a gravação podem ser automaticamente iniciadas em uma data específica e em um horário específico (por exemplo, segunda-feira, 20:00) e por uma duração de tempo específica (por exemplo, entre 20:00 e 21:00). Alternativamente, a monitoração e a gravação podem ser iniciadas em resposta a uma entrada de usuário (por exemplo, um clique de mouse, uma tecla de função ou combinação de tecla) a partir de um dispositivo de controle (por exemplo, um controle remoto, etc.). Em algumas implementações, o áudio ambiente é codificado usando-se uma variação de transmissão contínua dos recursos discriminativos de 32 bits/de quadro descritos em Ke et al.

Em algumas implementações, o software de cliente roda uma "barra lateral" ou um outro elemento de interface de usuário. Dessa forma, quando a interface de lado de cliente tem o boot dado, a amostragem de áudio ambiente pode começar imediatamente e rodar no "fundo" com resultados (opcionalmente) sendo exibidos na barra lateral, sem se invocar uma sessão de navegador da web completa.

Em algumas modalidades, a amostragem de áudio ambiente pode começar quando a interface de lado de cliente tem o boot dado ou quando o espectador faz um login em um serviço ou aplicativo (por exemplo, e-mail, etc.).

Os descritores são enviados para o servidor de banco de dados de áudio. Em algumas implementações, os descritores são sumários estatísticos comprimidos do áudio ambiente, conforme descrito em Ke et al. Pelo envio de sumários estatísticos, a privacidade acústica do usuário é mantida, porque os sumários estatísticos não são reversíveis, isto é, o áudio original não pode ser recuperado a partir do descritor. Assim, quaisquer conversações pelo usuário ou outros indivíduos monitorados e gravados no am-

biente de difusão não podem ser reproduzidas pelo descritor. Em algumas implementações, os descritores podem ser encriptados para privacidade extra e segurança, usando-se uma ou mais técnicas de encriptação (por exemplo, encriptação de chave assimétrica ou simétrica, encriptação elíptica, etc.).

Em algumas implementações, os descritores são enviados para o servidor de banco de dados de áudio 104 como uma submissão de consulta (também referido como descritor de consulta), em resposta a um evento de gatilho detectado pelo processo de monitoração na interface de lado de cliente 102. Por exemplo, um evento de gatilho poderia ser o tema de abertura de um programa de televisão (por exemplo, a melodia de abertura de "Seinfeld") ou um diálogo falado pelos atores. Em algumas implementações, os descritores de consulta podem ser enviados para o servidor de banco de dados de áudio 104 como parte do processo de transmissão contínua. Em algumas implementações, os descritores de consulta podem ser transmitidos para o servidor de banco de dados de áudio 104 em resposta a uma entrada de usuário (por exemplo, através de controle remoto, cliques de mouse, etc.).

#### Processo de Personalização de Massa

A Figura 3 é um fluxograma de um processo de personalização de massa 300. As etapas do processo 300 não têm que ser completadas em qualquer ordem em particular, e pelo menos algumas das etapas podem ser realizadas ao mesmo tempo em um ambiente de processamento de linha múltipla ou em paralelo.

O processo 300 começa quando uma interface de lado de cliente (por exemplo, a interface de lado de cliente 102) monitora e grava fragmentos de áudio ambiente de uma difusão de mídia de massa em um ambiente de difusão (302). Os fragmentos de áudio ambiente gravados são codificados em descritores (por exemplo, sumários estatísticos comprimidos), os quais podem ser enviados para um servidor de banco de dados de áudio (304) como consultas. O servidor de banco de dados de áudio compara as consultas em relação a um banco de dados de descritores de referência

computados a partir de estatísticas de difusão de mídia de massa para a determinação de descritores candidatos que mais bem combinem com a consulta (308). Os descritores candidatos são enviados para um servidor de aplicativo social ou um outro recurso de rede, o qual usa os descritores candidatos para a agregação da informação personalizada para o usuário (310). Por exemplo, se o usuário estiver assistindo ao show de televisão "Seinfeld", então, os descritores de consulta gerados a partir do áudio ambiente do show serão combinados com descritores de referência derivados de difusões prévias de "Seinfeld". Assim, os descritores candidatos de melhor combinação são usados para a agregação de uma informação personalizada relativa a "Seinfeld" (por exemplo, notícias, grupos de discussão, links para comunidades de par social *ad hoc* ou salas de bate-papo, anúncios, etc.). Em algumas implementações, o procedimento de combinação é eficientemente realizado usando-se técnicas de hashing (por exemplo, hashing direto ou hashing de sensibilidade a localidade (LSH)) para a obtenção de uma lista curta de descritores candidatos, conforme descrito com respeito à Figura 4. Os descritores candidatos então são processados em um procedimento de validação, tal como descrito em Ke et al.

Em algumas implementações, os descritores de consulta de espectadores diferentes são diretamente combinados, ao invés de se combinar cada consulta com um banco de dados de descritores de referência. Uma modalidade como essa permitiria a criação de comunidades de par social *ad hoc* sobre um assunto para o qual um banco de dados de descritores de referência não estivesse disponível. Uma modalidade como essa poderia combinar em tempo real espectadores que estivessem na mesma forma de público (por exemplo, estádio, bar, etc.) usando dispositivos eletrônicos portáteis (por exemplo, telefones móveis, PDAs, etc.).

#### Classificações de Popularidade

Em algumas implementações, estatísticas em tempo real e agregadas são inferidas a partir de uma lista de espectadores atualmente assistindo à difusão (por exemplo, show, anúncio, etc.). Estas estatísticas podem ser acumuladas no fundo, enquanto os espectadores estiverem usando

outros aplicativos. As estatísticas podem incluir, mas não estão limitadas a: 1) o número médio de espectadores assistindo à difusão; 2) o número médio de tempos em que os espectadores assistiram à difusão; 3) outros shows a que os espectadores assistiram; 4) o número mínimo e o de pico de espectadores; 5) para que os espectadores mais freqüentemente mudaram quando eles deixaram uma difusão; 6) por quanto tempo os espectadores assistiram a uma difusão; 7) quantas vezes os espectadores mudaram de canal; 8) quais anúncios foram assistidos pelos espectadores; e 9) a partir do que os espectadores mais freqüentemente mudaram quando entraram em uma difusão, etc. A partir destas estatísticas, uma ou mais classificações de popularidade podem ser determinadas.

As estatísticas usadas para gerar classificações de pluralidade podem ser geradas, usando-se um contador para cada canal de difusão sendo monitorado. Em algumas implementações, os contadores podem ser inseridos com dados de grupo demográfico ou dados de grupo geográfico. As classificações de popularidade podem ser usadas por espectadores para "verem o que é quente", enquanto a difusão estiver em andamento (por exemplo, ao notar uma classificação aumentada durante a performance do intervalo de meio tempo do Super Bowl de 2004). Os anunciantes e provedores de conteúdo também podem usar as classificações de popularidade para dinamicamente ajustarem o material descrito em resposta às classificações. Isto é especialmente verdadeiro para anúncios, uma vez que o comprimento unitário curto e numerosas versões de anúncios geradas por campanhas de propaganda são facilmente trocados para ajuste aos níveis de classificação de espectador. Outros exemplos de estatísticas incluem, mas não estão limitados a: popularidade de uma difusão de televisão versus uma difusão de rádio por demografia ou tempo, a popularidade de horários do dia, isto é, os horários de pico de assistência/audiência, o número de lares em uma dada área, a quantidade de zape de canal durante shows em particular (gênero de show, horários do dia em particular), o volume da difusão, etc.

A informação personalizada é enviada para a interface de lado de cliente (312). As classificações de popularidade também são armazena-

das em um banco de dados para uso por outros processos (318), tal como o ajuste dinâmico de anúncios descritos acima. A informação personalizada é recebida na interface de lado de cliente (314), onde é formatada e apresentada para uma interface de usuário (316). A informação personalizada pode ser associada a um meio de comentário (por exemplo, mensagens de texto em uma sala de bate-papo) que é apresentado ao usuário em uma interface de usuário. Em algumas implementações, uma sala de bate-papo pode incluir um ou mais subgrupos. Por exemplo, um grupo de discussão para "Seinfeld" poderia incluir um subgrupo denominado "Especialistas em Seinfeld" ou um subgrupo pode ser associado a uma demografia em particular, tais como mulheres entre as idades de 20 e 30 que assistem a "Seinfeld", etc.

Em algumas implementações, a informação bruta (por exemplo, valores de contador) usada para a geração de estatísticas para classificações de popularidade é coletada e armazenada na interface de lado de cliente, ao invés de no servidor de aplicativo social. A informação bruta pode ser transferida para a difusora sempre que o usuário estiver on-line e/ou invocar um aplicativo de personalização de massa.

Em algumas implementações, uma caixa de medição de difusão (BMB) é instalada na interface de lado de cliente. A BMB pode ser um dispositivo de hardware simples, que seja similar a uma set-top box, mas não se conecta ao dispositivo de difusão. Diferentemente do sistema de classificação de Nielsen, o qual requer que um hardware seja instalado na televisão, a BMB pode ser instalada próxima do sistema de mídia de massa ou dentro do alcance do sinal de televisão. Em algumas implementações, a BMB automaticamente grava fragmentos de áudio e gera descritores, os quais são armazenados em uma memória (por exemplo, uma memória flash). Em algumas implementações, a BMB opcionalmente pode incluir um ou mais botões de hardware os quais podem ser pressionados por um usuário para indicação de a qual difusão eles estão assistindo (similar às classificações de Nielsen). O dispositivo de BMB pode ser recolhido pelo provedor de classificações de tempos em tempos para a coleta dos descritores armazenados, ou a BMB pode difundir os descritores armazenados para uma ou mais partes interes-

sadas por uma conexão de rede (por exemplo, telefone, Internet, rádio sem fio, tal como um serviço de mensagem curta (SMS) para rádio, etc.) de tempos em tempos.

5 Em algumas implementações, os anúncios podem ser monitorados para a determinação da efetividade do anúncio, a qual pode ser reportada de volta para os anunciantes. Por exemplo, quais anúncios foram assistidos, pulados, nível de volume dos anúncios, etc.

10 Em algumas implementações, um dispositivo de captura de imagem (por exemplo, uma câmera digital, um gravador de vídeo, etc.) pode ser usado para a medição de quantos espectadores estão assistindo ou ouvindo a uma difusão. Por exemplo, vários algoritmos de combinação de padrão conhecidos podem ser aplicados a uma imagem ou a uma seqüência de imagens para a determinação do número de espectadores presentes em um ambiente de difusão, durante uma difusão em particular. As imagens e/ou os  
15 dados derivados das imagens pode(m) ser usado(s) em combinação com descritores de áudio para a acumulação de uma informação personalizada para um usuário, computação de classificações de popularidade ou para qualquer outra finalidade.

#### Processo de Impressão Digital de Áudio

20 A Figura 4 é um fluxograma de um processo de impressão digital de áudio 400. As etapas de processo 400 não têm que ser completadas em qualquer ordem em particular e pelo menos algumas etapas podem ser realizadas ao mesmo tempo em um ambiente de processamento de linha múltipla ou de processamento em paralelo. O processo 400 combina descritores  
25 de consulta gerados em uma interface de lado de cliente (por exemplo, a interface de lado de cliente 102) com descritores de referência armazenados em um ou mais bancos de dados em tempo real e com baixa latência. O processo 400 adapta uma técnica proposta por Ke et al. para a manipulação de dados de áudio ambiente (por exemplo, a partir de uma difusão de televi-  
30 são) e consultas.

O processo 400 começa em uma interface de lado de cliente pela decomposição de fragmentos de áudio ambiente (por exemplo, de 5 a 6

segundos de áudio) de uma difusão de mídia de massa capturada por um detector de áudio ambiente (por exemplo, um microfone) em quadros de superposição (por exemplo, espaçados 12 ms). Cada quadro é convertido em um descritor (por exemplo, um descritor de 32 bits) que é treinado para suplant

5 ruir ruído de áudio e distorção (404), conforme descrito em Ke et al. Em algumas implementações, cada descritor representa um sumário estatístico de identificação do fragmento de áudio.

Em algumas implementações, os descritores podem ser enviados como fragmentos de consulta (por exemplo, referidos como descritores de consulta) para um servidor de banco de dados de áudio em que eles são combinados com um banco de dados de descritores de referência identificando sumários estatísticos de fragmentos de áudio previamente gravados da difusão de mídia de massa (406). Uma lista de descritores candidatos tendo melhores combinações pode ser determinada (408). Os descritores candidatos podem ser classificados em escores, de modo que descritores candidatos que sejam temporalmente consistentes com o descritor de consulta tenham um escore mais alto do que descritores candidatos que sejam menos temporalmente consistentes com o descritor de consulta (410). Os descritores candidatos com os escores mais altos (por exemplo, um escore excede a um valor de limite suficientemente alto) são transmitidos ou de outra forma providos para um servidor de aplicativo social (412), onde eles podem ser usados para a agregação da informação personalizada relacionada à difusão de mídia. Usar um limite assegura que os descritores sejam suficientemente combinados, antes de os descritores serem transmitidos ou providos de outra forma para o servidor de aplicativo social (412).

10  
15  
20  
25

Em algumas implementações, o banco de dados de descritores de referência pode ser gerado a partir de difusões dadas por várias companhias de mídia, os quais podem ser indexados e usados para a geração dos descritores. Em outras implementações, descritores de referência também podem ser gerados usando-se guias de televisão ou outros metadados e/ou uma informação embutida no sinal de difusão.

30

Em algumas implementações, uma tecnologia de reconhecimento

to de fala pode ser usada para ajudar na identificação de qual programa está sendo assistido. Essa tecnologia poderia ajudar os usuários a discutirem novos eventos ao invés de apenas shows de televisão. Por exemplo, um usuário poderia assistir a um lançamento de vôo espacial em um canal diferente de um outro espectador e, portanto, possivelmente obter um sinal de áudio diferente (por exemplo, devido a um locutor diferente). Uma tecnologia de reconhecimento de fala poderia ser usada para o reconhecimento de palavras chaves (por exemplo, vôo espacial, lançamento, etc.), as quais poderiam ser usadas para se ligar o usuário a um meio de comentário.

#### 10 Descritores de Hashing

Ke et al. Usa técnicas de visão em computador para encontrar estatísticas compactas altamente discriminativas para áudio. Seu procedimento foi treinado em pares rotulados de exemplos positivos (em que  $x$  e  $x'$  são versões com ruído do mesmo áudio) e exemplos negativos (em que  $x$  e  $x'$  são de um áudio diferente). Durante esta fase de treinamento, uma técnica de aprendizado de máquina baseada em intensificação usa os pares rotulados para a seleção de uma combinação de 32 filtros e limites que conjuntamente criam uma estatística altamente discriminativa. Os filtros localizam mudanças na magnitude de espectrograma, usando diferenças de primeira e de segunda ordem através do tempo e da frequência. Um benefício do uso destes filtros de diferença simples é que eles podem ser calculados eficientemente usando-se uma técnica de imagem integral descrita em Viola, P. and Jones, M. (2002), *Robust Real-Time Object Detection*, *International Journal of Computer Vision*, o qual é incorporado aqui como referência em sua totalidade.

Em algumas implementações, as saídas destes 32 filtros são limites, dando um bit único por filtro em cada quadro de áudio. Estes resultados de 32 limites formam apenas descritores transmitidos daquele quadro de áudio. Esta escassez na codificação assegura a privacidade do usuário quanto a uma escuta não autorizada. Ainda, estes descritores de 32 bits são robustos para as distorções de áudio nos dados de treinamento, de modo que exemplos positivos (por exemplo, quadros de combinação) tenham dis-

tâncias de Hamming pequenas (isto é, número de bits de diferença de medição de distância) e exemplos negativos (por exemplo, quadros não combinados) tenham grandes distâncias de Hamming. Deve ser notado que mais ou menos filtros podem ser usados e mais de um bit pode filtro pode ser usado em cada quadro de áudio (por exemplo, mais bits usando múltiplos testes de limite).

Em algumas implementações, o descritor de 32 bits em si é usado como uma chave de hash para um hashing direto. O descritor é uma função de hash bem equilibrada. As taxas de recuperação são adicionalmente melhoradas pela consulta não apenas do descritor de consulta, mas também um conjunto pequeno de descritores similares (até uma distância de Hamming de 2 a partir do descritor de consulta original).

#### Consistência Temporal em Consulta

Uma vez que os descritores de consulta sejam combinados com o banco de dados de áudio usando-se o procedimento de hashing descrito acima, as combinações são validadas para se determinarem quais acertos de retorno de banco de dados são combinações acuradas. Caso contrário, um descritor candidato poderia ter muitos quadros combinados com o descritor de consulta, mas com a estrutura temporal errada.

Em algumas implementações, uma validação é obtida pela visualização de cada acerto de banco de dados como um suporte para uma combinação em um deslocamento de banco de dados de consulta específico. Por exemplo, se o descritor oito ( $q_8$ ) em um fluxograma de consulta de "Seinfeld" de 415 quadros de comprimento de 5 segundos,  $q$ , acertasse o 1008º descritor de banco de dados ( $x_{1008}$ ), este suportaria uma combinação de candidato entre a consulta de 5 segundos e os quadros 1001 a 1415 no banco de dados de áudio. Outras combinações entre  $q_n$  e  $x_{1000+n}$  ( $1 \leq n \leq 415$ ) suportariam esta mesma combinação de candidato.

Além de uma consistência temporal, precisa-se considerar quadros quando conversações temporariamente abafam o áudio ambiente. Isto pode ser modelado como uma comutação exclusiva entre áudio ambiente e sons interferentes. Para cada quadro de consulta  $i$ , há uma variável oculta,

$y_i$ : se  $y_i=0$ , o  $i$ ésimo quadro da consulta será modelado como interferência apenas; se  $y_i=1$ , o  $i$ ésimo quadro será modelado como a partir de um áudio ambiente limpo. Tomar uma visão extrema (ambiente puro ou interferência pura) é justificado pela precisão extremamente baixa com a qual cada quadro de áudio é representado (32 bits) e suavizado pela provisão de probabilidades de Bit-Flop de bit adicionais para cada uma das 32 posições do vetor de quadro sob cada uma das duas hipóteses ( $y_i=0$  e  $y_i=1$ ). Finalmente, são modelados as transições entre quadros entre estados apenas de ambiente e estados apenas de interferência como um processo de Markov de primeira ordem oculto, com probabilidades de transição derivadas a partir dos dados de treinamento. Por exemplo, pode-se reusar o modelo de probabilidade de 66 parâmetros dado por Ke et al., CVPR 2005.

O modelo final da probabilidade de combinação entre um vetor de consulta,  $q$ , e um vetor de banco de dados de ambiente em um deslocamento de  $N$  quadros,  $x_N$ , é:

$$P(q \mid x^N) = \prod_{n=1}^{415} P(\langle q_n, x_{N+n} \rangle \mid y_n) P(y_n \mid y_{n-1}) \quad (1)$$

onde  $\langle q_n, x_m \rangle$  denota as diferenças de bit entre os vetores de quadro de 32 bits  $q_n$  e  $x_m$ . Este modelo incorpora a restrição de consistência temporal e o modelo de Markov oculto de ambiente/interferência.

#### 20 Filtração de Consistência Pós-Combinação

As pessoas freqüentemente falam com outras enquanto assistem à televisão, resultando em uma interferência acústica esporádica, mas forte, especialmente quando se usam microfones baseados em laptop para a amostragem do áudio ambiente. Dado que a maioria das expressões vocais de conversação são de dois ou três segundos de duração, uma troca de comunicação simples entre espectadores tornaria uma consulta de 5 segundos irreconhecível.

Em algumas implementações, uma filtração pós-combinação é usada para se lidar com estas não combinações de confiança baixa intermitentes. Por exemplo, pode-se usar um modelo de Markov oculto de tempo contínuo de troca de canal com um tempo de espera esperado (isto é, um

tempo entre mudanças de canal) de  $L$  segundos. O servidor de aplicativo social 106 indica a combinação de confiança mais alta no passado recente (juntamente com sua confiança "descontada") como parte de uma informação de estado associada a cada sessão de cliente. Usando esta informação, o servidor 106 seleciona a combinação de índice de conteúdo a partir do passado recente ou a combinação de índice atual, com base naquela delas que tiver a confiança mais alta.

Foram usados  $M_h$  e  $C_h$  para uma referência à melhor combinação para o incremento de tempo prévio (5 segundos atrás) e seu escore de confiança de probabilidade em log. Se simplesmente foi aplicado o modelo de Markov a esta melhor combinação prévia, sem tomar uma outra observação, então, a expectativa é que a melhor combinação para o tempo atual seja a mesma seqüência de programa, exatamente de 5 segundos a mais de comprimento, e a confiança nesta expectativa é  $C_h - l/L$ , onde  $l = 5$  segundos é o incremento de consulta. Este desconto de  $l/L$  na probabilidade em log corresponde à probabilidade de modelo de Markov,  $e^{-l/L}$ , de não mudar canais durante o incremento de tempo de comprimento  $l$ .

Uma hipótese alternativa é gerada pela combinação de áudio para a consulta atual. Foi usado  $M_0$  para uma referência à melhor combinação para o fragmento de áudio atual: isto é a combinação que é gerada pelo processo de impressão digital de áudio 400.  $C_0$  é o escore de confiança de probabilidade em log dado pelo processo de impressão digital de áudio 400.

Se estas duas combinações (a expectativa histórica atualizada e a observação de fragmento atual) derem combinações diferentes, será selecionada a hipótese com o escore de confiança mais alto:

$$\{M_0, C_0\} = \begin{cases} \{M_h, C_h - l/L\} & \text{e } C_h - l/L > C_0 \\ M_0, C_0 & \text{Caso contrário} \end{cases} \quad (2)$$

(onde  $M_0$  é a combinação que é usada pelo servidor de aplicativo social 106 para a seleção de conteúdo relacionado e  $M_0$  e  $C_0$  são levados adiante no próximo incremento de tempo como  $M_h$  e  $C_h$ ).

### 30 Interface de Usuário

A Figura 5 é um fluxograma de uma modalidade de uma interfa-

ce de usuário 208 para interação com aplicativos de personalização de massa. A interface de usuário 208 inclui uma área de exibição de camada personalizada 502, uma área de exibição de meio de comentário 504, uma área de exibição de links patrocinados 506 e uma área de exibição de conteúdo 508. As camadas personalizadas podem ser navegadas usando-se uma barra de navegação 510 e um dispositivo de entrada (por exemplo, um mouse ou um controle remoto). Cada camada tem um rótulo associado na barra de navegação 510. Por exemplo, se o usuário selecionar o rótulo "Moda", então, a camada de moda, a qual inclui um conteúdo relacionado à moda associado a "Seinfeld", será apresentada na área de exibição 502.

Em algumas implementações, a interface de lado de cliente 102 inclui um dispositivo de exibição 210 capaz de apresentar uma interface de usuário 208. Em algumas implementações, a interface de usuário 208 é uma página da web interativa apresentada pelo servidor de aplicativo social 106 e apresentada em uma janela de navegador na tela do dispositivo de exibição 210. Em algumas implementações, a interface de usuário 208 é persistente e estará disponível para interação após o áudio de difusão usado no processo de combinação de conteúdo ter se deslocado no tempo. Em algumas implementações, a interface de usuário 208 é dinamicamente atualizada ao longo do tempo ou em resposta a um evento de disparo (por exemplo, uma nova pessoa entra na sala de bate-papo, um comercial começa, etc.). Por exemplo, a cada vez em que um comercial é difundido, a área de exibição de links patrocinados 506 pode ser atualizada com links novos 518 relacionados ao assunto do comercial.

Em algumas implementações, a informação personalizada e os links patrocinados podem ser enviados por e-mail para o espectador ou descritos em uma barra lateral em um tempo posterior.

Em algumas implementações, a interface de lado de cliente 102 recebe uma informação personalizada a partir do servidor de aplicativo social 106. Esta informação pode incluir uma página da web, um e-mail, um quadro de mensagens, links, mensagem instantânea, uma sala de bate-papo ou um convite para se unir a um grupo de discussão em aberto, eRoom, vi-

deconferência ou netmeeting, chamada de voz (por exemplo, pelo Skype®), etc. Em algumas implementações, a interface de usuário 208 provê acesso a comentários e/ou links para comentários a partir de difusões ou filmes previamente vistos. Por exemplo, se o usuário estiver atualmente assistindo a um DVD do "Shrek", ele poderá querer ver o que as pessoas disseram sobre o filme no passado.

Em algumas implementações, a área de exibição 502 inclui uma região de classificação 512, a qual é usada para a exibição de classificações de popularidade relacionadas a uma difusão. Por exemplo, a área de exibição 512 pode mostrar quantos espectadores estão atualmente assistindo a "Seinfeld", se comparado com um outro show que seja difundido ao mesmo tempo.

Em algumas implementações, a área de exibição de meio de comentário 504 apresenta um ambiente do tipo de sala de bate-papo em que múltiplos usuários podem comentar sobre difusões. Em algumas implementações, a área de exibição 504 inclui uma caixa de texto 514 para a introdução de comentários que são enviados para a sala de bate-papo usando-se o mecanismo de entrada 516 (por exemplo, um botão).

A área de exibição de links patrocinados 506 inclui uma informação, imagens e/ou links relacionados à propaganda que está associada à difusão. Por exemplo, um dos links 518 pode levar o usuário para um website que esteja vendendo mercadorias do "Seinfeld".

A área de exibição de conteúdo 508 é onde o conteúdo de difusão é exibido. Por exemplo, uma cena da difusão atual pode ser exibida com uma outra informação relevante (por exemplo, número de episódio, título, estampa de tempo, etc.). Em algumas implementações, a área de exibição 508 inclui controles 520 (por exemplo, botões de rolagem) para navegação através do conteúdo exibido.

#### Favoritos de Vídeo

Em algumas implementações, um botão 522 é incluído na área de exibição de conteúdo que pode ser usado para tornar favorito um vídeo. Por exemplo, ao clicar no botão 522, o episódio de "Seinfeld" descrito na

área de exibição 508 é adicionado à biblioteca de vídeo de favoritos do usuário, o qual então pode ser visto sob demanda através de um aplicativo de transmissão contínua baseado na web ou de outros métodos de acesso. De acordo com a política regulada pelo proprietário do conteúdo, o serviço de transmissão contínua pode prover uma reexecução de visualização única grátis, coletar pagamentos como o agente para os proprietários de conteúdo, ou inserir anúncios que proveriam pagamento para os proprietários de conteúdo.

### Arquitetura de Hardware de Interface de Lado de Cliente

10 A Figura 6 é um diagrama de blocos de uma arquitetura de hardware 600 para a interface de lado de cliente 102 descrita na Figura 1. Embora a arquitetura de hardware 600 seja típica de um dispositivo de computação (por exemplo, um computador pessoal), as implementações descritas podem ser realizadas em qualquer dispositivo capaz de apresentar uma interface de usuário em um dispositivo de exibição, incluindo, mas não limitando: computadores de mesa ou portáteis; dispositivos eletrônicos; telefones, telefones móveis; sistemas de exibição; televisores; monitores; sistemas de navegação; tocadores/gravadores de mídia portáteis; assistentes digitais pessoais; sistemas de jogos; dispositivos eletrônicos portáteis; e dispositivos ou aparatos eletrônicos embutidos.

20 Em algumas implementações, o sistema 600 inclui um ou mais processadores 602 (por exemplo, uma CPU), opcionalmente um ou mais dispositivos de exibição 604 (por exemplo, CRT, LCD, etc.), uma interface de microfone 606, uma ou mais interfaces de rede 608 (por exemplo, USB, Ethernet, portas FireWire®, etc.), opcionalmente um ou mais dispositivos de entrada 610 (por exemplo, mouse, teclado, etc.) e um ou mais meios que podem ser lidos em computador 612. Cada um destes componentes é operativamente acoplado a um ou mais barramentos 614 (por exemplo, EISA, PCI, USB, FireWire®, NuBus, PDS, etc.).

30 Em algumas implementações, não há dispositivos de exibição ou dispositivos de entrada, e o sistema 600 apenas realiza amostragem e codificação (por exemplo, gerando descritores, etc.) no fundo, sem uma entrada

de usuário.

O termo "meio que pode ser lido em computador" refere-se a qualquer meio que participe na provisão de instruções para um processador 602 para execução, incluindo, sem limitação, meios não-voláteis (por exemplo, discos óticos ou magnéticos), meios voláteis (por exemplo, uma memória) e meios de transmissão. A mídia de transmissão inclui, sem limitação, cabos coaxiais, fio de cobre e fibra ótica. Os meios de transmissão também podem assumir a forma de ondas acústicas, de luz ou frequência de rádio.

O(s) meio(s) que pode(m) ser lido(s) em computador inclui(em) um sistema operacional 616 (por exemplo, Mac OS®, Windows®, Unix, Linux, etc.), um módulo de comunicações de rede 618, um software de cliente 620 e um ou mais aplicativos 622. O sistema operacional 616 pode ser de usuário múltiplo, de processamento múltiplo, de multitarefa, de linha múltipla, em tempo real, e similares. O sistema operacional 616 realiza tarefas básicas, incluindo, mas não limitando: reconhecimento de uma entrada a partir de dispositivos de entrada 610; envio de saída para dispositivos de exibição 604; acompanhamento de arquivos e diretórios em dispositivos de armazenamento 612; controle de dispositivos periféricos (por exemplo, unidades de disco, impressoras, dispositivos de captura de imagem, etc.); e gerenciamento de tráfego em um ou mais barramentos 614.

O módulo de comunicações de rede 618 inclui vários componentes para o estabelecimento e a manutenção de conexões de rede (por exemplo, um software para a implementação de protocolos de comunicações, tais como TCP/IP, HTTP, Ethernet, USB, FireWire®, etc.).

O software de cliente 620 provê vários componentes de software para implementação do lado de cliente dos aplicativos de personalização de massa e para a realização das várias funções de lado de cliente descritas com respeito às Figura 1 a 5 (por exemplo, identificação de áudio ambiente). Em algumas implementações, alguns ou todos os processos realizados pelo software de cliente 620 podem ser integrados no sistema operacional 616. Em algumas implementações, o processo pode ser pelo menos parcialmente implementado em um circuito eletrônico digital, ou em um hardware, firmwa-

re ou software de computador ou em qualquer combinação dos mesmos.

Outros aplicativos 624 podem incluir qualquer outro aplicativo de software, incluindo, mas não limitando: processadores de texto, navegadores, e-mail, Envio de Mensagem Instantânea, tocadores de mídia, software de telefonia, etc.

### Detecção de Anúncios e Redifusões

#### Detecção de Repetição

Quando da preparação de um banco de dados para uma busca, ajuda ser capaz de pré-indicar um material repetido usando os descritores descritos previamente. O material de repetição pode incluir, mas não está limitado à repetição de shows, anúncios, subsegmentos (por exemplo, imagens de arquivo em shows de notícias), etc. Usando-se estes indicadores, um material repetido pode ser apresentado de uma forma que não empurre todo o outro material além do vão de atenção de um usuário conduzindo uma busca (por exemplo, além dos 10 a 20 primeiros acertos). O processo 700 descrito abaixo provê uma forma de detecção daquelas duplicatas antes de quaisquer consultas de busca no banco de dados.

#### Remoção de Anúncio de Vídeo

Uma das reclamações das difusoras tinha a ver com permitir que o material a ser buscado e tocado fosse a redifusão de uma propaganda embutida. Do ponto de vista das difusoras, esta redifusão é contraproducente: ela diminui o valor das difusões que o anunciante paga diretamente, uma vez que provê ao anunciante uma propaganda grátis. A menos que os anúncios antigos sejam removidos e novos anúncios sejam postos no lugar de uma forma que retorne algum repasse para as difusoras originais, elas não tiram lucro da reexibição de seu material difundido previamente. O processo 700 descrito abaixo provê uma forma de detecção de anúncio embutido ao procurar por repetições, possivelmente em conjunto com outros critérios (por exemplo, duração, volume, atividade visual, delineação de quadros em branco, etc.).

#### Sumarização de Vídeo

Se um "sumário" (isto é, versão mais curta) de um material de

programa não repetido for necessário, uma forma de se obter isso é remover os anúncios (conforme detectado pelo material repetido) e tomar segmentos a partir do material imediatamente precedente e imediatamente seguinte à localização do anúncio. Em uma difusão de televisão, estas posições no programa tipicamente contêm "enigmas" (antes dos anúncios) e "recapitulações" (imediatamente após os anúncios). Se um sumário for para ser feito de um programa de notícias que inclui uma mistura de material não de propaganda não repetido e repetido, tipicamente, o material não de propaganda repetido corresponderá a uma declaração curta. Estes segmentos geralmente contribuem com menos informação do que a narração do âncora da história do noticiário e são bons candidatos para remoção. Se um sumário for para ser feito de um programa narrativo (por exemplo, um filme ou um capítulo de série), trilhas de áudio repetidas tipicamente corresponde aos sons de tema, música incidental ou silêncio. De novo, estes tipicamente são bons segmentos para remoção de um vídeo de sumário. O processo 700 descrito abaixo provê uma forma de detecção dessas trilhas de áudio repetidas, de modo que elas possam ser removidas do vídeo de sumário.

#### Processo de Detecção de Repetição

A Figura 7 é um fluxograma de uma modalidade de um processo de detecção de repetição 700 de modo conforme. As etapas de processo 700 não têm que ser completadas em qualquer ordem em particular e pelo menos algumas etapas podem ser realizadas ao mesmo tempo em um ambiente de processamento de linha múltipla ou em paralelo.

O processo 700 começa pela criação de um banco de dados de estatísticas de áudio a partir de um conjunto de conteúdo, tais como alimentações de televisão, transferências (via upload) de vídeo, etc. (702). Por exemplo, o banco de dados poderia conter descritores de 32 bits/de quadro, conforme descrito em Ke et al. As consultas são feitas a partir do banco de dados e rodadas em relação ao banco de dados para se ver onde as repetições ocorrem (704). Em algumas implementações, um segmento curto de estatísticas de áudio é tomado como uma consulta e uma rodada checada quanto a combinações não de identidade (combinações que não sejam idên-

5 ticas) usando técnicas de hashing (por exemplo, hashing direto ou hashing de sensibilidade de localidade (LSH)) para a obtenção de uma lista curta de combinações auditivas possíveis. Estas combinações candidatas então são processadas em um procedimento de validação, por exemplo, conforme descrito em Ke et al. Um conteúdo correspondente a uma combinação de candidato validado pode ser identificado como um conteúdo de repetição (706).

10 As combinações não-identidade são as mais fortes são "feitas crescer" para frente e para trás no tempo, para se encontrar os pontos de começo e de fim do material repetido (708). Em algumas implementações, isto pode ser feito usando-se técnicas de programação dinâmica (por exemplo, decodificação de Viterbi). Na extensão da combinação para frente no tempo, a última fração de tempo na combinação de "origem" forte é regulada como a "combinação" e a última fração de tempo da primeira combinação de 15 intensidade crível abaixo para o mesmo deslocamento de banco de dados entre a consulta e a combinação é regulada como "não de combinação". Em algumas implementações, os escores de combinação para quadros individuais entre estes dois pontos fixos são usados como observações, e um modelo de Markov de primeira ordem permitindo transições em estado, mais uma 20 transição única de estados "de combinação" para "não de combinação", é usado. A probabilidade de transição de combinação para não combinação para  $1/L$  pode ser regulada de forma um pouco arbitrária, onde  $L$  é o número de quadros entre estes dois pontos fixos, correspondente ao conhecimento mínimo da localização de transição na faixa admitida. Uma outra possibilidade 25 de para a seleção de probabilidades de transição seria usar os perfis de intensidade de combinação para orientação desta estimativa para uma transição anterior ou posterior. Mas isto aumentaria a complexidade do modelo de programação dinâmico e não tem probabilidade de melhorar os resultados, uma vez que as probabilidades de combinação já são usadas como obser- 30 vações neste período. O mesmo processo é usado para crescimento de combinações de segmento para trás no tempo (por exemplo, apenas comutar para passado/futuro e rodar o mesmo algoritmo).

Em algumas implementações, as anotações de áudio são combinadas com uma informação não auditiva (por exemplo, anotações visuais) para a obtenção de acurácias de combinação mais altas. Por exemplo, as combinações que são encontradas com uma combinação de áudio então podem ser verificadas (ou checadas uma segunda vez) pelo uso de medidas de similaridade visual simples (710). Estas medidas podem incluir, mas não estão limitadas a: histogramas de cor (por exemplo, freqüências de cores similares em duas imagens), estatísticas sobre o número e a distribuição de bordas, etc. Estas não precisam ser computadas apenas pela imagem inteira, mas podem ser computadas para sub-regiões das imagens também e comparadas com as sub-regiões correspondentes na imagem alvo.

Para aqueles aplicativos que estejam procurando anúncios (em contraste com todos os tipos de material repetido), os resultados de detecção de material repetido podem ser combinados com medidas direcionadas para a distinção de anúncios de não anúncios (712). Estas características de distinção podem se basear em convenções de anúncio, tais como durações (por exemplo, spots de 10/15/30 segundos são comuns), em volume (por exemplo, os anúncios tendem a ser mais altos do que o material de programa circundante, de modo que se o material repetido for mais alto do que o material em um dos lados, é mais provável que seja um anúncio), em atividade visual (por exemplo, anúncios tendem a ter transições mais rápidas entre cenas e mais movimento em cena, de modo que se o material repetido tiver diferenças de quadro maiores do que o material em um dos lados, seja mais provável que seja um anúncio) e em delineação de quadros em branco (anúncios localmente inseridos tipicamente não preenchem completamente o intervalo que é deixado para eles pela alimentação nacional, resultando em quadros em branco e um silêncio em um espaçamento que é um múltiplo de 30 segundos).

Uma vez que os anúncios sejam identificados, o material circundando os anúncios pode ser analisado e estatísticas podem ser geradas. Por exemplo, as estatísticas podem ser geradas sobre quantas vezes um produto em particular é anunciado usando-se uma criação em particular (por e-

xemplo, imagens, texto) ou quantas vezes um segmento em particular vai ao ar, etc. Em algumas implementações, um ou mais anúncios antigos podem ser removidos ou substituídos por novos anúncios. Técnicas adicionais para detecção e substituição de anúncio são descritas em Covell, M., Baluja, S., Fink, M., Advertisement Detection and Replacement Using Acoustic and Visual Repetition, IEEE Signal Processing Society, MMSP 2006 International Workshop on Multimedia Signal Processing, October 3-6, 2006, BC Canada, cujo artigo é incorporado como referência aqui em sua totalidade.

Em algumas implementações, uma informação de proprietários de conteúdo sobre a estrutura detalhada do conteúdo (por exemplo, onde o material de anúncio foi inserido, onde os programas foram repetidos) poderia ser usada para se aumentar o processo 700 e elevar as acurácias de combinação. Em algumas implementações, estatísticas de vídeo podem ser usadas para a determinação da repetição, ao invés de áudio. Em outras implementações, uma combinação de estatísticas de vídeo e de áudio pode ser usada.

#### Leilões de Fragmento de Áudio

Em algumas implementações, os anunciantes podem participar em leilões relacionados à presença de áudio ambiente que esteja relacionado ao produto ou serviço que o anunciante quer vender. Por exemplo, múltiplos anunciantes poderiam ofertar em um leilão o direito de associação de seus produtos ou serviços a um fragmento de áudio ou descritor associado a "Seinfeld". O vencedor do leilão poderia então colocar alguma informação relacionada na frente do espectador (por exemplo, os links patrocinados), sempre que o áudio ambiente em questão estiver presente. Em algumas implementações, os anunciantes poderiam ofertar fragmentos de áudio ambiente tendo uma descrição de metanível. Por exemplo, os anunciantes poderiam ofertar um áudio que estivesse associado a um anúncio de televisão (por exemplo, este é o áudio associado ao anúncio de TV do Ford Explorer), em *closed caption* (por exemplo, a captação dizendo "Yankees baseball"), em uma localização de segmento de programa (por exemplo, este áudio ocorrerá em 15 min para o "Seinfeld" e ocorrerá 3 minutos após o intervalo

comercial prévio e 1 min antes do próximo intervalo comercial), ou em propriedades acústicas ou visuais de nível baixo (por exemplo, "música de fundo", "vozes em conversação", "tipo explosivo", etc.).

Em algumas implementações, um ou mais aplicativos de personalização podem ser rodados no fundo, enquanto o usuário realiza outras tarefas, tais como navegação por um outro website (por exemplo, um link patrocinado). O material que está relacionado a uma difusão de mistura (por exemplo, conteúdo de televisão) pode participar nos mesmos leilões de link patrocinado como um material que esteja relacionado a uma outra fonte de conteúdo (por exemplo, um conteúdo de website). Por exemplo, anúncios relacionados à TV podem ser misturados com anúncios que correspondam ao conteúdo de uma página da web atual.

Várias modificações podem ser feitas nas implementações descritas e ainda se estar no escopo das reivindicações a seguir.

## REIVINDICAÇÕES

1. Método, que compreende:  
a geração de uma consulta a partir de um banco de dados de estatísticas de áudio;  
5 a rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade; e se a combinação não de identidade existir,  
a identificação do conteúdo correspondente à consulta combinada como conteúdo de repetição.
- 10 2. Método, de acordo com a reivindicação 1, que ainda compreende:  
a validação da combinação não de identidade usando-se uma informação não auditiva.
- 15 3. Método, de acordo com a reivindicação 1, que ainda compreende:  
a determinação de pontos finais do conteúdo de repetição.
4. Método, de acordo com a reivindicação 3, onde os pontos finais são determinados usando-se técnicas de programação dinâmica.
- 20 5. Método, de acordo com a reivindicação 1, que ainda compreende:  
a aplicação de medidas ao conteúdo de repetição para se determinar se o conteúdo de repetição é um anúncio.
- 25 6. Método, de acordo com a reivindicação 5, onde as medidas são a partir de um grupo de medidas consistindo em duração de tempo, volume, atividade visual e delineação de quadro em branco.
7. Método, de acordo com a reivindicação 1, onde as estatísticas de áudio são geradas a partir de fragmentos de áudio ambiente de uma difusão de mídia.
- 30 8. Método, de acordo com a reivindicação 1, onde as estatísticas de áudio são descritores de quadro.
9. Método, de acordo com a reivindicação 1, onde as estatísticas de vídeo são usadas com as estatísticas de áudio para a determinação de

uma combinação não de identidade.

10. Método, de acordo com a reivindicação 1, onde as combinações não de identidade são determinadas usando-se técnicas de hashing.

11. Sistema, que compreende:

5 um processador;

um meio que pode ser lido em computador operativamente acoplado ao processador e tendo instruções armazenadas ali, as quais, quando executadas pelo processador, fazem com que o processador realize as operações de:

10 geração de uma consulta a partir de um banco de dados de estatísticas de áudio;

rodada de uma consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade, onde as estatísticas de áudio são geradas a partir de conteúdo; e

15 se uma combinação não de identidade for encontrada, a identificação do conteúdo correspondente à consulta combinada como um conteúdo de repetição.

12. Sistema, de acordo com a reivindicação 11, onde o processo ainda realiza a operação de:

20 validação da combinação não de identidade usando-se uma informação não auditiva.

13. Sistema, de acordo com a reivindicação 11, onde o processador ainda realiza a operação de:

determinação de pontos finais do conteúdo de repetição.

25 14. Sistema, de acordo com a reivindicação 13, onde os pontos finais são determinados usando-se técnicas de programação dinâmica.

15. Sistema, de acordo com a reivindicação 11, onde o processador ainda realiza a operação de:

30 aplicação de medidas ao conteúdo de repetição para se determinar se o conteúdo de repetição é um anúncio.

16. Sistema, de acordo com a reivindicação 15, onde as medidas são a partir de um grupo de medidas consistindo em duração de tempo,

volume, atividade visual e delimitação de quadro em branco.

17. Sistema, de acordo com a reivindicação 11, onde as estatísticas de áudio são geradas a partir de fragmentos de áudio ambiente de uma difusão de mídia.

5 18. Sistema, de acordo com a reivindicação 11, onde as estatísticas de áudio são descritores de quadro.

19. Sistema, de acordo com a reivindicação 11, onde as estatísticas de vídeo são usadas com as estatísticas de áudio para a determinação de uma combinação não de identidade.

10 20. Sistema, de acordo com a reivindicação 11, onde as combinações não de identidade são determinadas usando-se técnicas de hashing.

21. Sistema, que compreende:

um meio para a geração de um banco de dados de estatísticas de áudio a partir de um conteúdo;

15 um meio para a geração de uma consulta a partir de um banco de dados de estatísticas de áudio;

um meio para a rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade; e

20 se a combinação não de identidade existir,  
um meio para a identificação do conteúdo correspondente à consulta combinada como conteúdo de repetição.

22. Meio que pode ser lido em computador, que tem instruções armazenadas nele, as quais, quando executadas pelo processador, fazem com que o processador realize as operações de:

25 geração de uma consulta a partir de um banco de dados de estatísticas de áudio;

rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade; e

30 se a combinação não de identidade existir,  
identificação do conteúdo correspondente à consulta combinada como conteúdo de repetição.

23. Método, que compreende:

a geração de um banco de dados de estatísticas de áudio ambiente associadas a uma difusão de mídia;

a geração de uma consulta a partir do banco de dados;

5 a rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de combinações não de identidade;

a identificação de conteúdo de repetição com base em combinações positivas entre a consulta e o banco de dados de estatísticas de áudio;

a determinação de pontos finais do conteúdo de repetição;

10 a identificação de conteúdo antes e depois dos pontos finais do conteúdo de repetição; e

a geração de estatísticas com base no conteúdo identificado.

24. Método, de acordo com a reivindicação 23, que ainda compreende:

15 a aplicação de pelo menos uma medida ao conteúdo de repetição para se determinar se o conteúdo de repetição é um anúncio.

25. Método, de acordo com a reivindicação 24, onde a medida está associada a uma extensão de difusão de mídia.

20 26. Método, de acordo com a reivindicação 24, onde a medida está associada a um volume da difusão de mídia.

SISTEMA DE PERSONALIZAÇÃO DE MASSA  
**100**

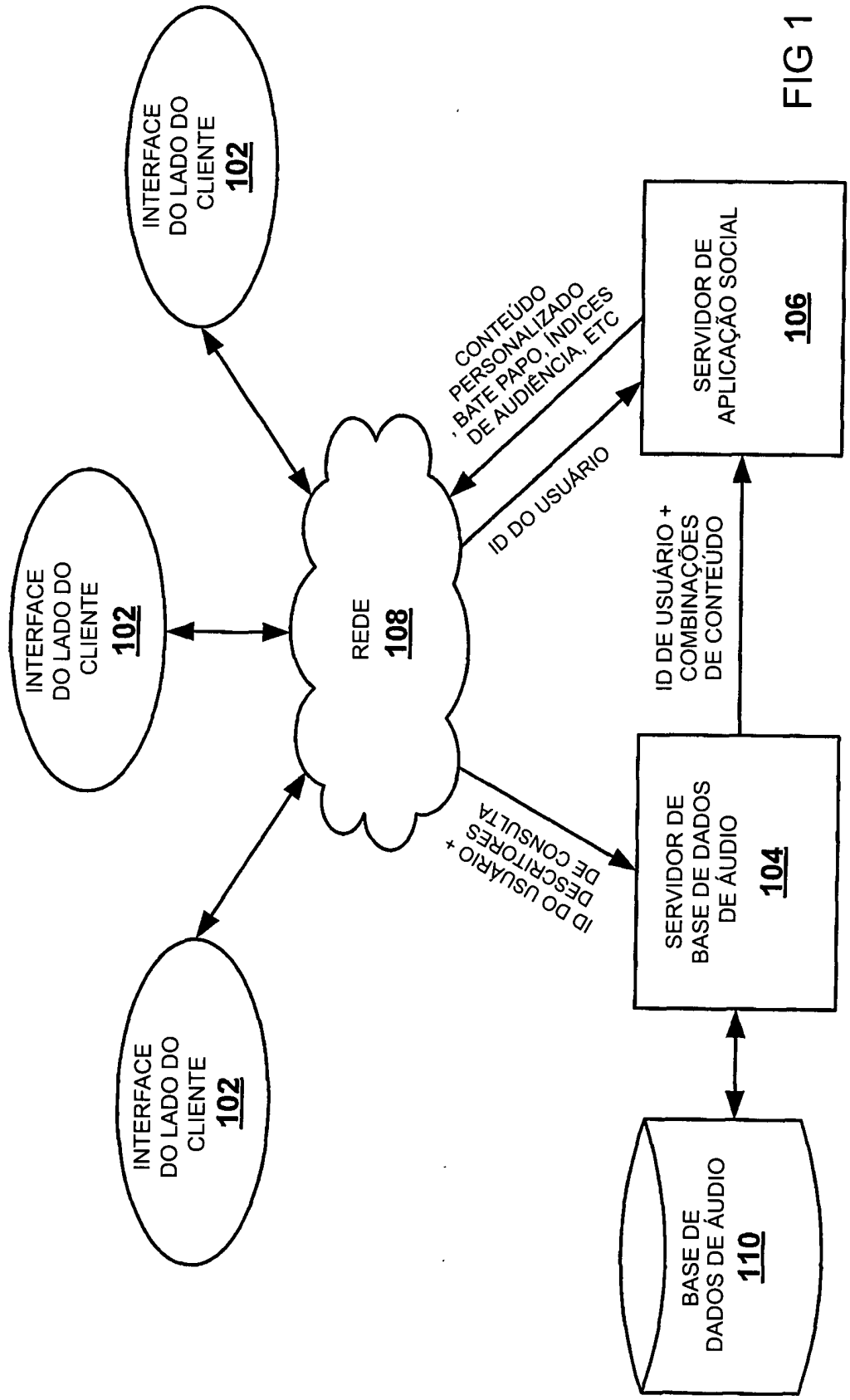


FIG 1

SISTEMA DE IDENTIFICAÇÃO DE ÁUDIO AMBIENTE (LADO DO CLIENTE)  
200

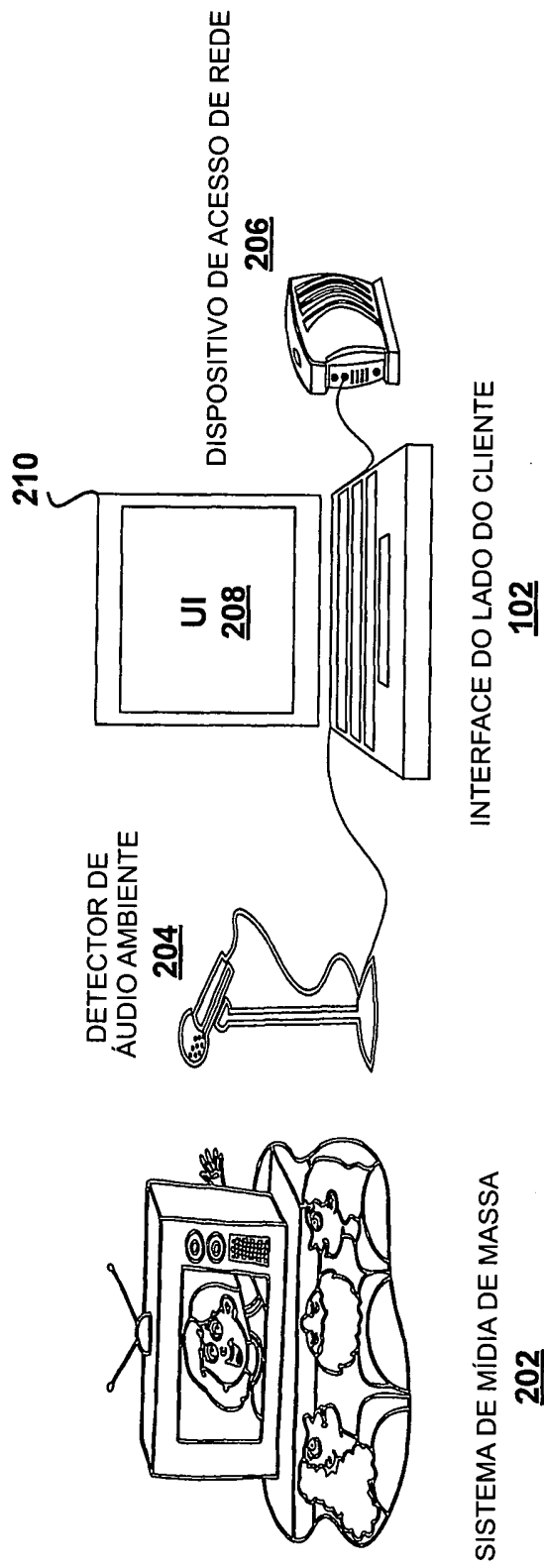


FIG 2

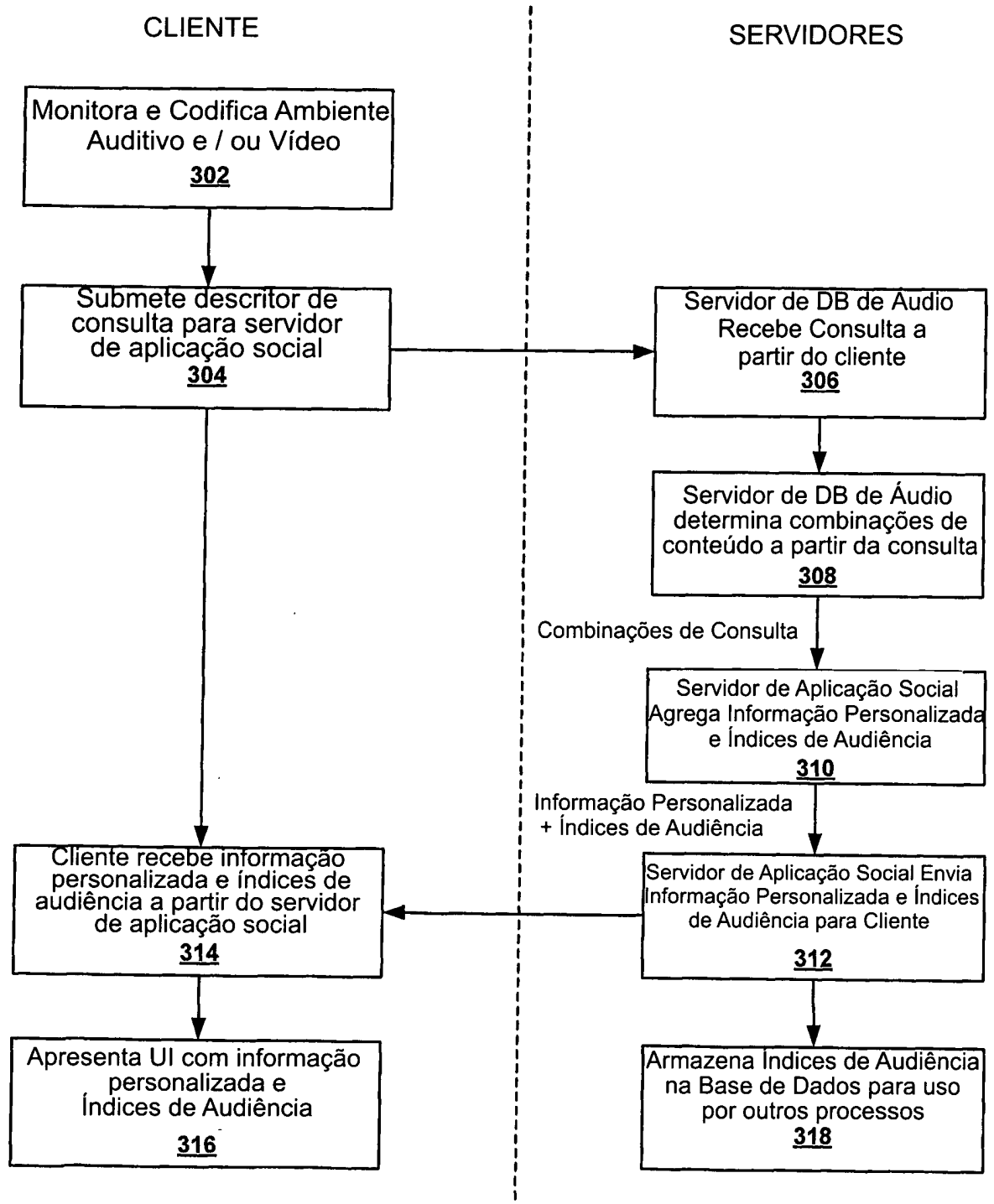


FIG 3

## Processo de Marcação de Impressão Digital de Áudio

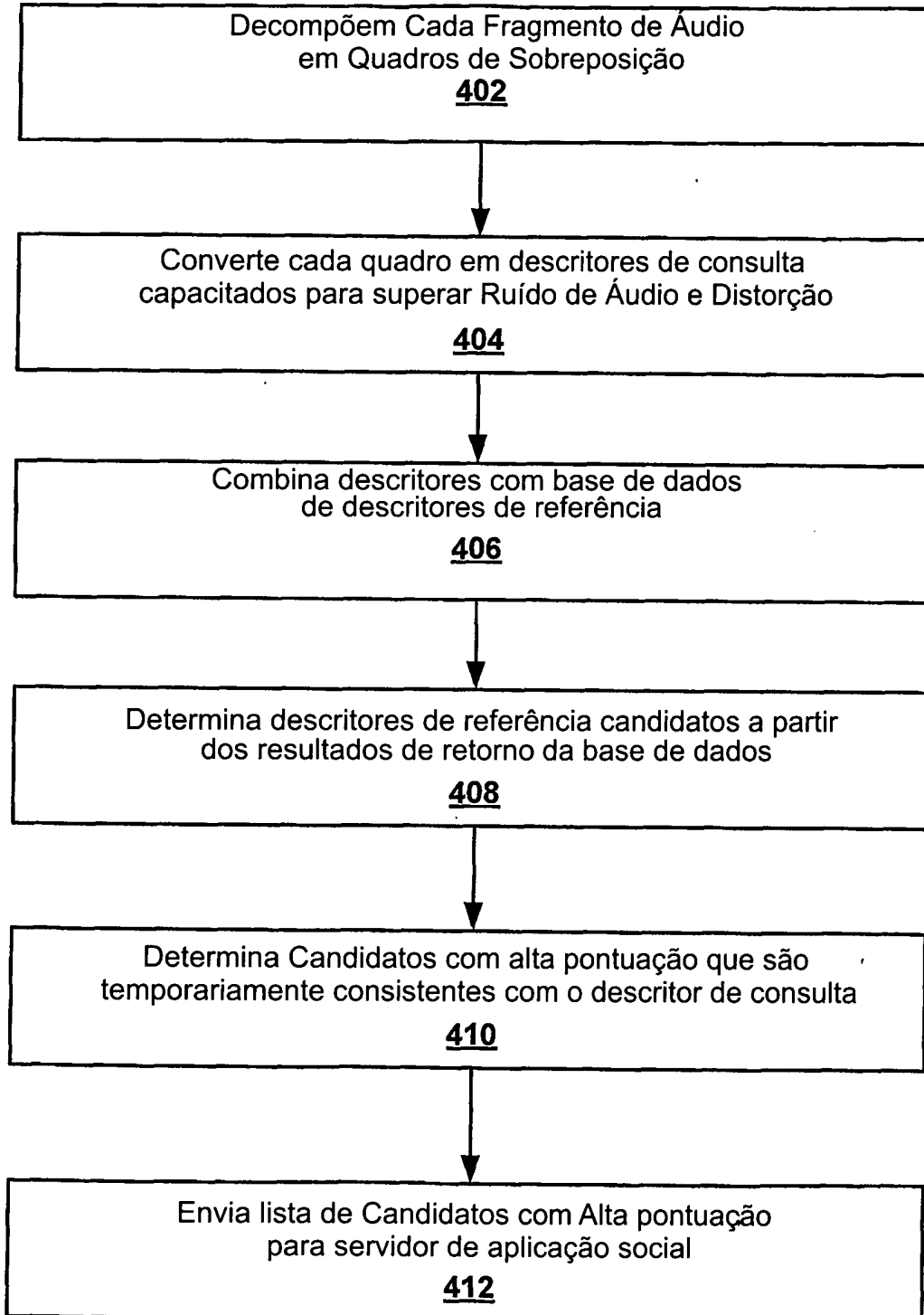
**400**

FIG 4

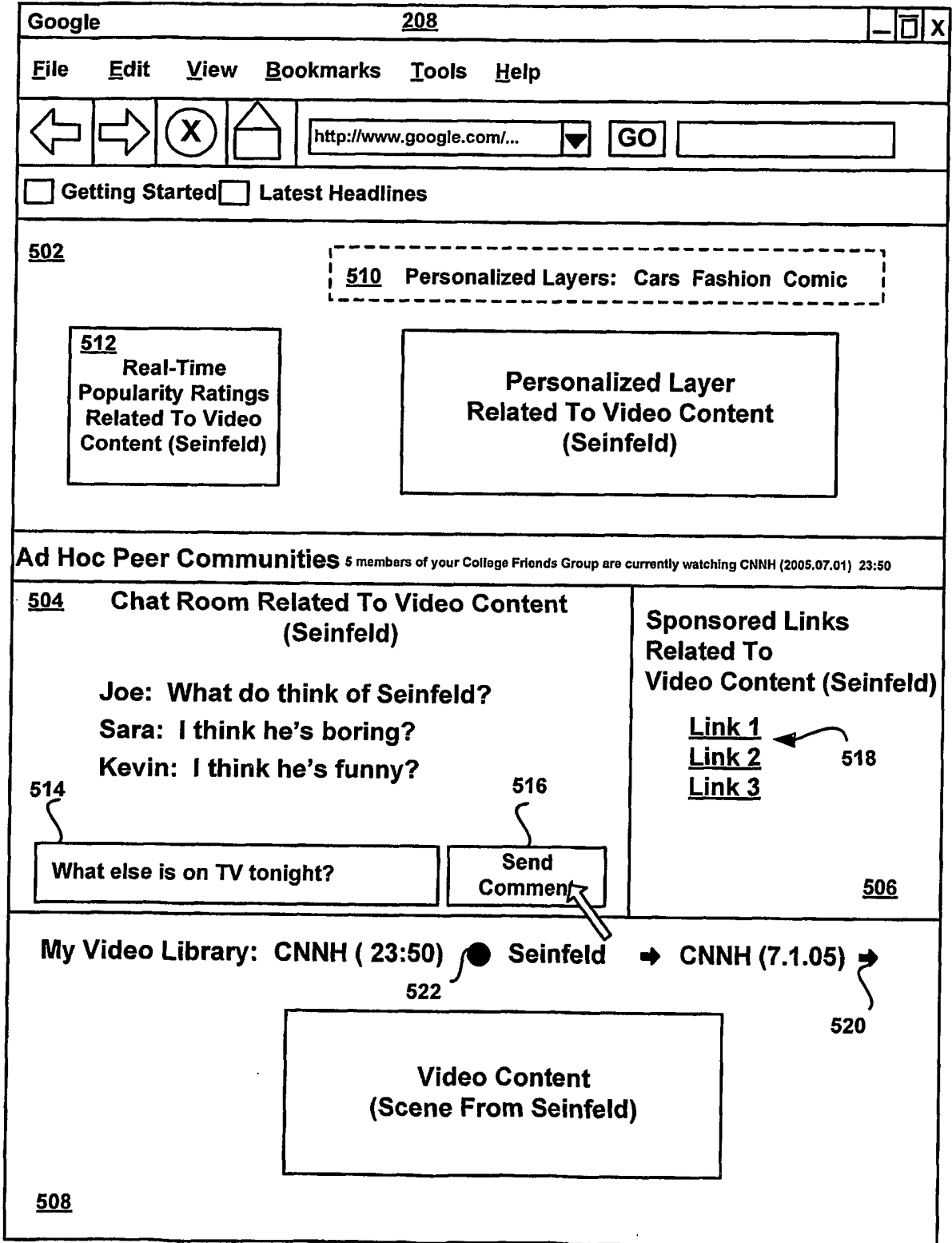


FIG 5

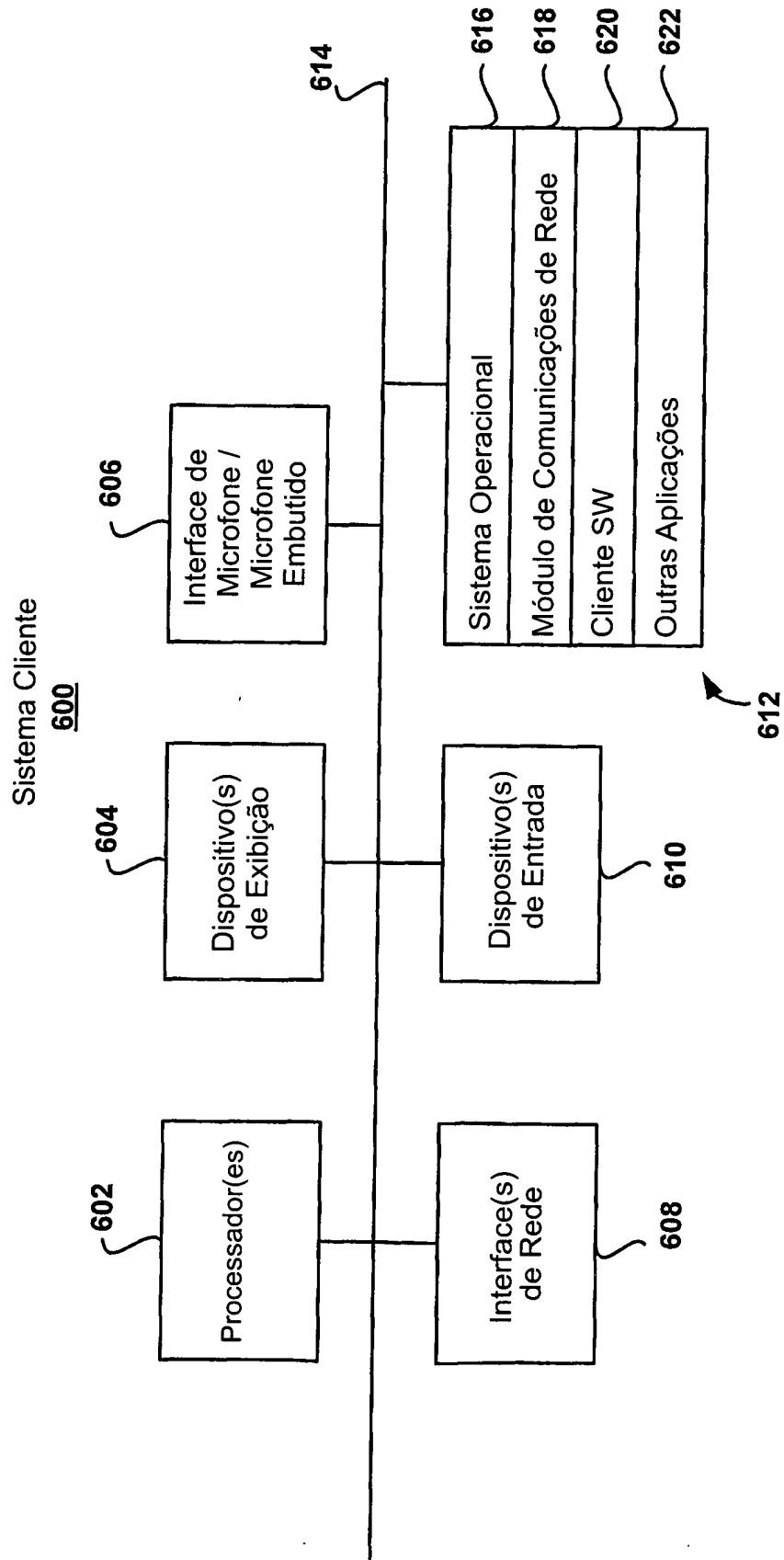


FIG 6

## Processo de Detecção de Repetição

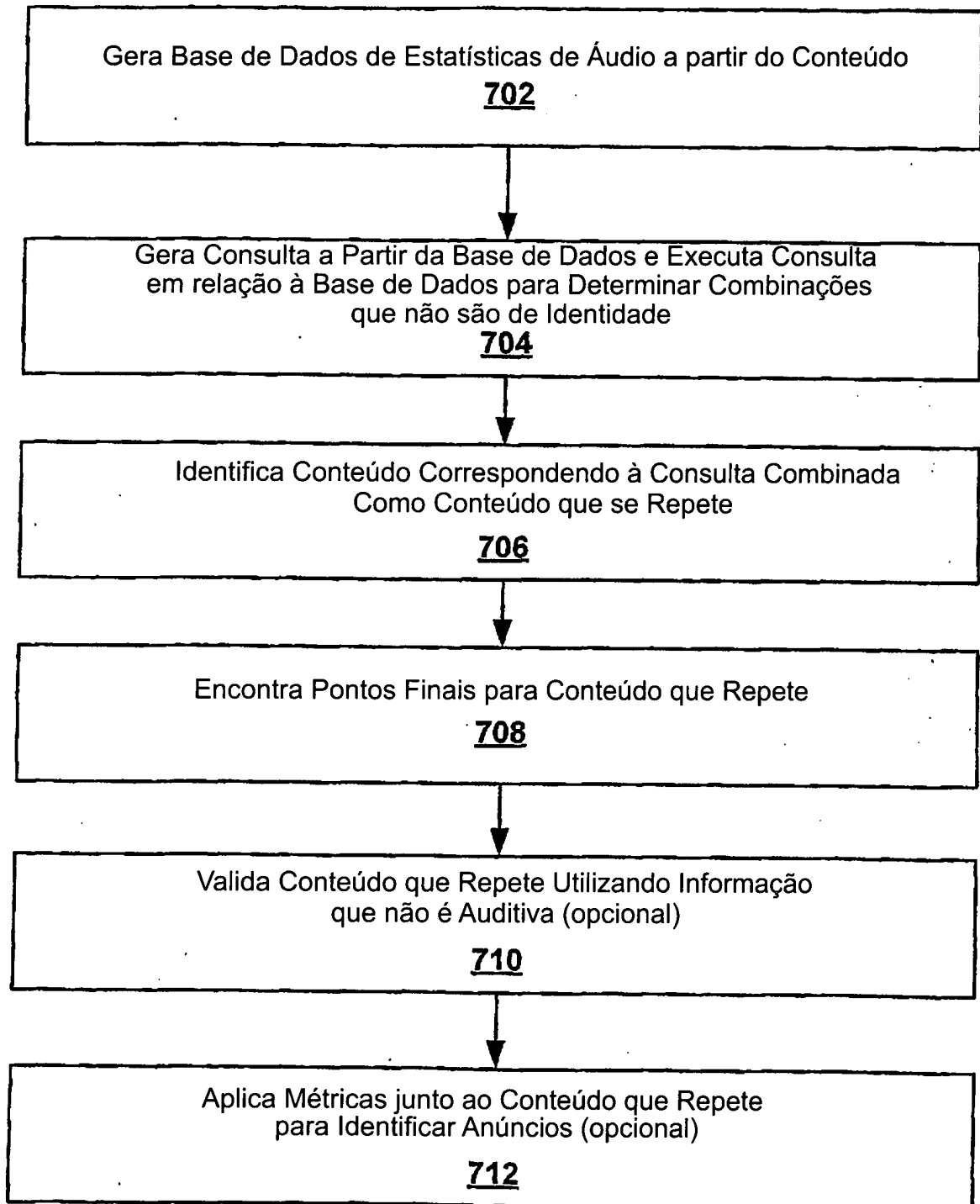
**700**

FIG 7

**RESUMO**

Patente de Invenção: **"DETECÇÃO DE CONTEÚDO DE REPETIÇÃO EM MÍDIA DE DIFUSÃO"**.

A presente invenção refere-se aos sistemas, métodos, dispositivos e produtos de programa de computador provêem aplicações sociais e interativas para a detecção de conteúdo de repetição em uma mídia de difusão. Em algumas implementações, um método inclui: a geração de um banco de dados de estatísticas de áudio a partir de um conteúdo; a geração de uma consulta a partir do banco de dados de estatísticas de áudio; a rodada da consulta em relação ao banco de dados de estatísticas de áudio para a determinação de uma combinação não de identidade; se uma combinação não de identidade existir, a identificação de conteúdo correspondente à consulta combinada como um conteúdo de repetição.