



(12) 发明专利

(10) 授权公告号 CN 101652807 B

(45) 授权公告日 2012. 09. 26

(21) 申请号 200880010930. 8

(22) 申请日 2008. 02. 01

(30) 优先权数据

60/887, 738 2007. 02. 01 US

(85) PCT申请进入国家阶段日

2009. 09. 29

(86) PCT申请的申请数据

PCT/US2008/052859 2008. 02. 01

(87) PCT申请的公布数据

W02008/095190 EN 2008. 08. 07

(73) 专利权人 缪斯亚米有限公司

地址 美国新泽西

(72) 发明人 R·D·陶布 J·A·卡巴尼拉

(74) 专利代理机构 永新专利商标代理有限公司

72002

代理人 钟胜光 王英

(51) Int. Cl.

G10H 1/00 (2006. 01)

(56) 对比文件

CN 1106949 A, 1995. 08. 16, 全文.

Paul M. Brossier et al. Fast labeling of notes in music signals. 《5th International Conference on Music Information Retrieval (ISMIR 04)》. 2004, 摘要、第 1-3 节、图 1.

Paul Brossier et al. Real-time temporal segmentation of note objects in music signals. 《Proceedings of ICMC 2004: The 30th Annual International Computer Music Conference》. 2004, 第 2.3 节.

Paul M. Brossier et al. Fast labeling of notes in music signals. 《5th International Conference on Music Information Retrieval (ISMIR 04)》. 2004, 摘要、第 1-3 节、图 1.

审查员 张鑫

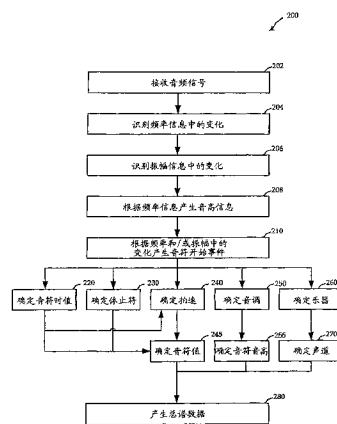
权利要求书 7 页 说明书 23 页 附图 16 页

(54) 发明名称

音乐转录的方法、系统及装置

(57) 摘要

本发明描述了用于将音频输入信号 (202) 数据自动转换成音乐总谱表示数据的方法、系统和设备。本发明的实施例识别来自音频信号的频率信息中的超过第一阈值的变化 (204) ;识别来自音频信号的振幅信息中的超过第二阈值的变化 (206) ;以及产生音符开始事件 (210), 每个音符开始事件表示音频信号中的识别出的频率信息中的超过第一阈值的变化和识别出的振幅信息中的超过第二阈值的变化中的至少一个的时间位置。音符开始事件和来自音频输入信号的其他信息的产生可用于提取音符音高 (255)、音符值 (245)、拍速 (240)、节拍、音调 (250)、乐器 (260) 和其他总谱表示信息。



CN 101652807 B

1. 一种用于从音频信号中产生总谱数据的系统,所述系统包括:  
音频接收器,其可操作来处理所述音频信号;以及  
音符识别单元,其可操作来接收经处理的音频信号,并响应于下列操作中的至少一项来产生与所述经处理的音频信号中的时间位置相关的音符开始事件:  
识别超过第一阈值的频率变化;以及  
识别超过第二阈值的振幅变化。
2. 如权利要求 1 所述的系统,其中所述音符识别单元包括:  
信号处理器,其包括:  
频率检测器单元,其可操作来识别所述音频信号中超过所述第一阈值的频率变化;以  
及  
振幅检测器单元,其可操作来识别所述音频信号中超过所述第二阈值的振幅变化;以  
及  
音符处理器,其包括音符开始事件发生器,所述音符开始事件发生器与所述频率检测器单元和所述振幅检测器单元操作性地通信,并可操作来产生所述音符开始事件。
3. 如权利要求 2 所述的系统,其中所述音符处理器还包括:  
第一包络发生器,其可操作来根据所述经处理的音频信号的幅度产生第一包络信号;  
第二包络发生器,其可操作来根据所述第一包络信号的平均功率值产生第二包络信号;以及  
控制信号发生器,其可操作来响应于所述第一包络信号中从第一方向到第二方向的变化而产生控制信号,使得所述变化在大于预定的控制时间的持续时间内延续;  
其中,所述振幅检测器单元响应于所述控制信号的幅度具有大于所述第二包络信号的幅度的值而识别所述音频信号的超过所述第二阈值的幅度变化。
4. 如权利要求 3 所述的系统,其中产生音符开始事件包括指示音频输入信号中相应于所述音符开始事件的时间戳值。
5. 如权利要求 4 所述的系统,其中第一包络函数包括近似于所述音频输入信号的在每个时间戳值处的幅度的函数,以及第二包络函数包括近似于所述第一包络函数的在平均间隔上的平均功率的函数。
6. 如权利要求 5 所述的系统,其中在每个时间戳值处的控制信号值被设置成等于所述第一包络函数在先前时间戳值处的最大幅度值,以及,响应于在一时间戳值处的第一包络函数值与对于大于第三阈值的时间间隔在值上不同的、先前时间戳值处的第一包络函数值之间的值的差异,将在所述时间戳值处的所述控制信号值改变为相比于先前控制信号值的负值。
7. 如权利要求 5 所述的系统,其中产生音符开始事件还包括响应于接收到的调节值来调节所述第二包络函数的所述平均间隔。
8. 如权利要求 7 所述的系统,其中所述接收到的调节值是根据从用户输入接收的乐器类别选择来确定的。
9. 如权利要求 7 所述的系统,其中所述接收到的调节值是根据从用户输入接收的音乐种类选择来确定的。
10. 如权利要求 2 所述的系统,还包括:

音符时值检测器单元,其与所述音符开始事件发生器操作性地通信,并可操作来至少通过确定第一音符开始事件和第二音符开始事件之间的时间间隔来检测音符时值,所述第一音符开始事件和所述第二音符开始事件由所述音符开始事件发生器产生,所述第二音符开始事件在时间上在所述第一音符开始事件之后;以及

将所述音符时值与所述第一音符开始事件关联起来,其中所述音符时值表示所确定的时间间隔。

11. 如权利要求 6 所述的系统,还包括:

音符时值检测器单元,其与所述音符开始事件发生器操作性地通信,并可操作来至少通过确定第一音符开始事件和第二音符开始事件之间的时间间隔来检测音符时值,所述第一音符开始事件和所述第二音符开始事件由所述音符开始事件发生器产生,所述第二音符开始事件在时间上在所述第一音符开始事件之后;以及

将所述音符时值与所述第一音符开始事件关联起来,其中所述音符时值表示所确定的时间间隔,

其中所述阈值是相应于作为音符时值的函数的时间间隔的可调节的值。

12. 如权利要求 10 所述的系统,其中所述第二音符开始事件是在时间上在所述第一音符开始事件之后的最接近的音符开始事件。

13. 如权利要求 3 所述的系统,还包括:

音符结束事件检测器单元,其可操作来在所述控制信号的振幅变得小于所述第二包络信号的振幅时产生与所述音频信号中的时间位置相关的音符结束事件;以及

音符时值检测器单元,其与所述音符开始事件发生器和所述音符结束事件检测器单元操作性地通信,并可操作来:

至少通过确定音符开始事件和音符结束事件之间的时间间隔来检测音符时值,所述音符结束事件在时间上在所述音符开始事件之后;以及

将所述音符时值与所述音符开始事件关联起来,其中所述音符时值表示所确定的时间间隔。

14. 如权利要求 1 所述的系统,还包括:

休止符检测器单元,其可操作来通过识别所述音频信号中振幅低于休止符检测阈值的部分来检测休止符。

15. 如权利要求 14 所述的系统,其中所述休止符检测器还可操作来通过确定音高置信度值小于音高置信度阈值来检测休止符,其中所述音高置信度值表示所述音频信号的所述部分包括与音频开始事件有关的音高的可能性。

16. 如权利要求 1 所述的系统,还包括:

拍速检测器单元,其与所述振幅检测器单元操作性地通信,并可操作来通过执行如下步骤来产生一组拍速数据:

确定一组参考拍速;

确定一组参考音符时值,每个参考音符时值表示预定的音符类型以每个参考拍速持续的时间长度;

确定拍速提取窗,所述拍速提取窗表示所述音频信号从第一时间位置延续到第二时间位置的连续部分;

通过确定所述音符开始事件出现在所述音频信号的所述连续部分内的位置来产生一组音符开始事件；

针对每个音符开始事件产生音符间隙，每个音符间隙表示在所述一组音符开始事件中的该音符开始事件和下一随后的音符开始事件之间的时间间隔；

产生一组误差值，每个误差值与相关的参考拍速关联起来，其中产生所述一组误差值的所述步骤包括：

将每个音符间隙除以所述一组参考音符时值中的每个参考音符时值；

将所述除法步骤中的每个结果四舍五入到在所述除法步骤中使用的参考音符时值的最接近的倍数；以及

评估在所述四舍五入步骤中的每个结果和所述除法步骤中的每个结果之间的差的绝对值；

识别所述一组误差值中的最小误差值；以及

确定与所述拍速提取窗相关的提取的拍速，其中所述提取的拍速是与所述最小误差值关联的相关参考拍速。

17. 如权利要求 16 所述的系统，其中所述拍速检测器单元还可操作来：

确定一组第二参考音符时值，每个参考音符时值表示一组预定的音符类型中的每一个音符类型以所述提取的拍速持续的时间长度；

针对每个音符开始事件，产生接收到的音符时值；以及

针对每个接收到的音符时值，确定接收到的音符值，所述接收到的音符值表示最佳地近似于所述接收到的音符时值的第二参考音符时值。

18. 如权利要求 2 所述的系统，还包括：

音调检测器单元，其与所述频率检测器单元操作性地通信，并可操作来通过执行如下步骤以产生一组音调数据：

确定一组代价函数，每个代价函数与一音调关联并表示一组预定频率中的每个频率对所关联的音调的符合度；

确定音调提取窗，所述音调提取窗表示所述音频信号从第一时间位置延续到第二时间位置的连续部分；

通过确定所述音符开始事件出现在所述音频信号的所述连续部分内的位置来产生一组音符开始事件；

确定所述一组音符开始事件中的每个音符开始事件的音符频率；

基于对照所述一组代价函数中的每一个代价函数评估所述音符频率来产生一组音调误差值；以及

确定接收到的音调，其中所述接收到的音调是与产生最低音调误差值的代价函数关联的音调。

19. 如权利要求 18 所述的系统，其中所述音调检测器单元还可操作来：

产生一组参考音高，每个参考音高表示一组预定音高中的一预定音高和所述接收到的音调之间的关系；以及

确定每个音符开始事件的音调音高标志，所述音调音高标志表示最佳地近似于该音符开始事件的音符频率的参考音高。

20. 如权利要求 2 所述的系统,还包括:

音色检测器单元,其与所述频率检测器单元操作性地通信,并可操作来检测与音符开始事件相关的音色数据。

21. 如权利要求 20 所述的系统,还包括:

声道检测器单元,其与所述音色检测器单元和所述频率检测器单元操作性地通信,并可操作来通过执行如下步骤以检测存在于所述音频信号中的音频声道:

产生一组音符开始事件,每个音符开始事件以至少一组音符特征为特点,所述一组音符特征包括音符频率和音符音色;

识别存在于所述音频信号中的多个音频声道,每个音频声道以一组声道特征为特点,所述一组声道特征包括音高映射或音色映射中的至少一个;以及

为每个音符开始事件的每组音符特征分配假定的声道,所述假定的声道是以最接近地匹配该组音符特征的一组声道特征为特点的音频声道。

22. 如权利要求 2 所述的系统,还包括:

包络检测器单元,其与所述振幅检测器单元操作性地通信,并可操作来确定与音符开始事件的起声、衰减、维持或释放中的至少一个有关的一组包络信息。

23. 如权利要求 20 所述的系统,还包括:

乐器识别单元,其与所述音色检测器单元操作性地通信,并可操作来至少部分地基于所述音色数据与音色样本的数据库的比较来识别乐器,每个音色样本与一乐器类型有关。

24. 如权利要求 20 所述的系统,还包括:

乐器识别单元,其包括与所述音色检测器单元操作性地通信的神经网络,所述神经网络可操作来至少部分地基于对照预定的代价函数评估所述音色数据来识别乐器。

25. 如权利要求 22 所述的系统,还包括:

乐器识别单元,其与所述包络检测器单元操作性地通信,并可操作来至少部分地基于所述包络信息与包络样本的数据库的比较来识别乐器,每个包络样本与一乐器类型有关。

26. 如权利要求 16 所述的系统,还包括:

节拍检测器单元,其与所述拍速检测器单元操作性地通信,并可操作来至少部分地通过使用神经网络对照一组节拍代价函数评估所述一组拍速数据来确定所述音频信号在节拍检测窗期间出现的部分的节拍。

27. 如权利要求 26 所述的系统,其中所述一组节拍代价函数与振幅信息或音高信息中的至少一个有关。

28. 如权利要求 1 所述的系统,其中所述音频信号包括具有与音乐表演有关的信息的数字信号。

29. 如权利要求 1 所述的系统,其中所述音频信号接收自一个或多个音频源,每个音频源选自自由麦克风、数字音频部件、音频文件、声卡和媒体播放器组成的组。

30. 一种用于从音频信号产生总谱数据的方法,所述方法包括:

识别来自所述音频信号的频率信息中的超过第一阈值的变化;

识别来自所述音频信号的振幅信息中的超过第二阈值的变化;以及

产生音符开始事件,每个音符开始事件表示所述音频信号中的、所识别出的频率信息中的超过所述第一阈值的变化和所识别出的振幅信息中的超过所述第二阈值的变化中的

至少一个的时间位置。

31. 如权利要求 30 所述的方法,还包括:

将音符记录与所述音符开始事件关联起来,所述音符记录包括一组音符特征数据。

32. 如权利要求 31 所述的方法,其中所述一组音符特征数据包括音高、振幅、包络、时间戳、时值或置信度中的至少一个。

33. 如权利要求 30 所述的方法,还包括:

产生第一包络信号,其中所述第一包络信号实质上追踪所述音频信号的振幅信息的绝对值;

产生第二包络信号,其中所述第二包络信号实质上追踪所述第一包络信号的平均功率;以及

产生控制信号,其中所述控制信号实质上追踪在所述第一包络信号中持续得比预定控制时间长方向变化的;

其中识别振幅信息中的变化的所述步骤包括识别表示所述音频信号中的一时间位置的第一音符开始位置,其中在该时间位置,所述控制信号的振幅变得大于所述第二包络信号的振幅。

34. 如权利要求 33 所述的方法,其中产生音符开始事件的所述步骤包括指示音频输入信号中相应于所述音符开始事件的时间戳值。

35. 如权利要求 34 所述的方法,其中第一包络函数包括近似于所述音频输入信号在每个时间戳值处的幅度的函数,以及第二包络函数包括近似于所述第一包络函数的在平均间隔上的平均功率的函数。

36. 如权利要求 35 所述的方法,其中在每个时间戳值处的控制信号值被设置成等于所述第一包络函数在先前时间戳值处的最大幅度值,以及,响应于在一时间戳值处的第一包络函数值与对于大于第三阈值的时间间隔在值上不同的、先前时间戳值处的第一包络函数值之间的值的差异,将在所述时间戳值处的所述控制信号值改变为相比于先前控制信号值的负值。

37. 如权利要求 35 所述的方法,其中产生音符开始事件的所述步骤还包括响应于接收到的调节值来调节所述第二包络函数的所述平均间隔。

38. 如权利要求 37 所述的方法,其中所述接收到的调节值根据从用户输入接收的乐器类别来确定。

39. 如权利要求 37 所述的方法,其中所述接收到的调节值根据从用户输入接收的音乐种类选择来确定。

40. 如权利要求 33 所述的方法,还包括:

识别表示所述音频信号中的一时间位置的第二音符开始位置,其中在该时间位置,所述控制信号的振幅在所述第一时间位置之后第一次变得大于所述第二包络信号的振幅;以及

将一时值与所述音符开始事件关联起来,其中所述时值表示从所述第一音符开始位置到所述第二音符开始位置的时间间隔。

41. 如权利要求 33 所述的方法,还包括:

识别表示所述音频信号中的一时间位置的音符结束位置,其中在该时间位置,所述控

制信号的振幅在所述第一音符开始位置之后第一次变得小于所述第二包络信号的振幅；以及

将一时值与所述音符开始事件关联起来，其中所述时值表示从所述第一音符开始位置到所述音符结束位置的时间间隔。

42. 如权利要求 36 所述的方法，还包括：

将一时值与所述音符开始事件关联起来，

其中，所述第三阈值是相应于作为音符时值的函数的时间间隔的可调节的值。

43. 如权利要求 30 所述的方法，还包括：

通过识别所述音频信号中振幅低于休止符检测阈值的部分来检测休止符。

44. 如权利要求 43 所述的方法，其中检测休止符的所述步骤还包括：确定小于音高置信度阈值的音高置信度值，其中所述音高置信度值表示所述音频信号的所述部分包括与音符开始事件有关的音高的可能性。

45. 如权利要求 30 所述的方法，还包括：

确定一组参考拍速；

确定一组参考音符时值，每个参考音符时值表示预定的音符类型以每个参考拍速持续的时间长度；

确定拍速提取窗，所述拍速提取窗表示所述音频信号从第一时间位置延续到第二时间位置的连续部分；

通过确定所述音符开始事件出现在所述音频信号的所述连续部分内的位置来产生一组音符开始事件；

针对每个音符开始事件产生音符间隙，每个音符间隙表示在所述一组音符开始事件中的该音符开始事件和下一随后的音符开始事件之间的时间间隔；

产生一组误差值，每个误差值与相关的参考拍速关联起来，其中产生所述一组误差值的所述步骤包括：

将每个音符间隙除以所述一组参考音符时值中的每个参考音符时值；

将所述除法步骤中的每个结果四舍五入到在所述除法步骤中使用的参考音符时值的最接近的倍数；以及

评估在所述四舍五入步骤中的每个结果和所述除法步骤中的每个结果之间的差的绝对值；

识别所述一组误差值中的最小误差值；以及

确定与所述拍速提取窗相关的提取的拍速，其中所述提取的拍速是与所述最小误差值关联的相关参考拍速。

46. 如权利要求 45 所述的方法，还包括：

确定一组第二参考音符时值，每个参考音符时值表示一组预定的音符类型中的每一个音符类型以所述提取的拍速持续的时间长度；

针对每个音符开始事件，产生接收到的音符时值；以及

针对每个接收到的音符时值，确定接收到的音符值，所述接收到的音符值表示最佳地近似于所述接收到的音符时值的第二参考音符时值。

47. 如权利要求 30 所述的方法，还包括：

确定一组代价函数,每个代价函数与一音调关联并表示一组预定频率中的每个频率对所关联的音调的符合度;

确定音调提取窗,所述音调提取窗表示所述音频信号从第一时间位置延续到第二时间位置的连续部分;

通过确定所述音符开始事件出现在所述音频信号的所述连续部分内的位置来产生一组音符开始事件;

确定所述一组音符开始事件中的每个音符开始事件的音符频率;

基于对照所述一组代价函数中的每一个代价函数评估所述音符频率来产生一组音调误差值;以及

确定接收到的音调,其中所述接收到的音调是与产生最低音调误差值的代价函数关联的音调。

48. 如权利要求 47 所述的方法,还包括:

产生一组参考音高,每个参考音高表示一组预定音高中的一个预定音高和所述接收到的音调之间的关系;以及

确定每个音符开始事件的音调音高标志,所述音调音高标志表示最佳地近似于该音符开始事件的音符频率的参考音高。

49. 如权利要求 30 所述的方法,还包括:

产生一组音符开始事件,每个音符开始事件以至少一组音符特征为特点,所述一组音符特征包括音符频率和音符音色;

识别存在于所述音频信号中的多个音频声道,每个音频声道以一组声道特征为特点,所述一组声道特征包括音高映射或音色映射中的至少一个;以及

为每个音符开始事件的每组音符特征分配假定的声道,所述假定的声道是以最接近地匹配该组音符特征的一组声道特征为特点的音频声道。

50. 一种用于从音频信号产生总谱数据的装置,所述装置包括:

用于识别来自所述音频信号的频率信息中的超过第一阈值的变化的模块;

用于识别来自所述音频信号的振幅信息中的超过第二阈值的变化的模块;以及

用于产生音符开始事件的模块,每个音符开始事件表示所述音频信号中的、所识别出的频率信息中的超过所述第一阈值的变化或所识别出的振幅信息中的超过所述第二阈值的变化中的至少一个的时间位置。



## 音乐转录的方法、系统及装置

### 技术领域

[0001] 本申请要求 2007 年 2 月 1 日提交的标题为“MUSICTRANSCRIPTION”（律师签号 026287-000200US）的共同未决的美国临时专利申请号 60/887, 738 的优先权，该专利由此通过引用被并入，如在本文件中为了所有的目的充分阐述的。

### 背景技术

[0002] 本发明通常涉及音频应用，并且尤其是涉及音频分解和总谱生成。

[0003] 提供原始音频输入信号到总谱 (score) 数据的精确的实时转换用于转录 (transcription) 可能是合乎需要的。例如，音乐表演者（例如，使用歌唱和 / 或其他乐器现场或被录制）可能希望自动转录表演，以产生活页乐谱或将表演转换成可编辑的数字总谱文件。很多元素可为音乐表演的部分，包括音符、音色 (timbre)、调式、力度强弱、节奏和声道 (track)。表演者可能要求所有这些元素都可从音频文件被可靠地提取，以产生准确的总谱。

[0004] 传统系统通常在这些领域中只提供有限的的能力，并且即使这些能力通常也是以有限的准确性和时间性来提供输出的。例如，很多传统系统要求用户向系统提供数据（而不是音频信号），以帮助该系统将音频信号转换成有用的总谱数据。一个因而产生的限制是，向系统提供数据而不是原始音频信号可能是耗费时间或不合需要的。另一因而产生的限制是，用户可能对系统所需要的数据知道得不多（例如，用户可能不熟悉音乐理论）。又一因而产生的限制是，该系统可能必须提供广泛的用户接口能力，以允许将所需数据提供到系统（例如，系统可能必须有键盘、显示器等）。

[0005] 因此，提供从原始音频文件自动和准确地提取总谱数据的提高的能力可能是合乎需要的。

### 发明内容

[0006] 描述了用于从音频信号自动和准确地提取总谱数据的方法、系统和设备。来自音频输入信号的频率信息中超过第一阈值的变化被识别出，以及来自音频输入信号的振幅信息中超过第二阈值的变化被识别出。产生音符开始事件 (note onset event)，使得每个音符开始事件表示音频输入信号中的、识别出的频率信息中超过第一阈值的变化或识别出的振幅信息中超过第二阈值的变化中的至少一个的时间位置。这里描述的技术可在具有包括在其中的计算机可读程序的计算机可读存储介质、方法和系统中实现。

[0007] 在本发明的一个方面，从一个或多个音频源接收音频信号。音频信号被处理以提取频率和振幅信息。频率和振幅信息用于检测音符开始事件（即，确定音符开始的时间位置）。对于每个音符开始事件，产生包络数据、音色数据、音高数据、力度强弱数据和其他数据。通过检查来自音符开始事件组的数据，产生拍速 (tempo) 数据、节拍 (meter) 数据、音调 (key) 数据、全局力度强弱数据、乐器和声道数据以及其他数据。各种数据接着用于产生总谱输出。

[0008] 在又一方面,从音频信号产生拍速数据,并确定一组参考拍速。确定一组参考音符时值 (note duration),每个参考音符时值表示预定音符类型以每个参考拍速持续的时间长度,并确定拍速提取窗,其表示音频信号从第一时间位置延续到第二时间位置的连续部分。一组音符开始事件通过如下操作产生:确定出现在音频信号的连续部分内的音符开始事件的位置;针对每个音符开始事件产生音符间隙 (note spacing),每个音符间隙表示该组音符开始事件中的音符开始事件和下一随后的音符开始事件之间的时间间隔;产生一组误差值,每个误差值与相关的参考拍速关联,其中产生该组误差值的步骤包括通过将每个音符间隙除以该组参考音符时值中的每个音符时值,将除法步骤的每个结果四舍五入到在除法步骤中使用的参考音符时值的最接近的倍数,以及算出四舍五入步骤的每个结果和除法步骤的每个结果之间的差的绝对值;识别该组误差值的最小误差值;以及确定与拍速提取窗相关的提取的拍速,其中所述提取的拍速是与最小误差值关联的相关参考拍速。拍速数据可进一步通过以下操作产生:确定一组第二参考音符时值,每个参考音符时值表示一组预定音符类型中的每一个以提取的拍速持续的时间长度;为每个音符开始事件产生接收到的音符时值;以及为每个接收到的音符时值确定接收到的音符值,该接收到的音符值表示最佳地近似于接收到的音符时值的第二参考音符时值。

[0009] 在又一方面,用于从音频信号产生音调数据的技术包括:确定一组代价函数,每个代价函数与音调关联并表示一组预定频率中的每个频率对相关音调的符合度;确定音调提取窗,其表示音频信号从第一时间位置延续到第二时间位置的连续部分;通过确定出现在音频信号的连续部分内的音符开始事件的位置来产生一组音符开始事件;为该组音符开始事件中的每个确定音符频率;基于对照该组代价函数中的每个算出音符频率来产生一组音调误差值;以及确定接收到的音调,其中接收到的音调是与产生最低音调误差值的代价函数关联的音调。在一些实施例中,该方法还包括:产生一组参考音高,每个参考音高表示一组预定音高中的一个预定音高和接收到的音调之间的关系;以及确定每个音符开始事件的音调音高标志 (key pitch designation),音调音高标志表示最佳地近似于音符开始事件的音符频率的参考音高。

[0010] 在又一方面,用于从音频信号产生声道数据的技术包括:产生一组音符开始事件,每个音符开始事件以至少一组音符特征为特点,该组音符特征包括音符频率和音符音色;识别存在于音频信号中的多个音频声道,每个音频声道以一组声道特征为特点,该组声道特征包括音高映射或音色映射中的至少一个;以及为每个音符开始事件分配每组音符特征的假定声道,假定声道是以最接近地匹配该组音符特征的一组声道特征为特点的音频声道。

[0011] 本发明的一个方面涉及一种用于从音频信号中产生总谱数据的系统,所述系统包括:音频接收器,其可操作来处理所述音频信号;以及音符识别单元,其可操作来接收经处理的音频信号,并响应于下列操作中的至少一项来产生与所述经处理的音频信号中的时间位置相关的音符开始事件:识别超过第一阈值的频率变化;以及识别超过第二阈值的振幅变化。

[0012] 本发明的另一个方面涉及一种用于从音频信号产生总谱数据的方法,所述方法包括:识别来自所述音频信号的频率信息中的超过第一阈值的变化;识别来自所述音频信号的振幅信息中的超过第二阈值的变化;以及产生音符开始事件,每个音符开始事件表示所

述音频信号中的、所识别出的频率信息中的超过所述第一阈值的变化和所识别出的振幅信息中的超过所述第二阈值的变化中的至少一个的时间位置。

[0013] 本发明的再一个方面涉及一种用于从音频信号产生总谱数据的装置,所述装置包括:用于识别来自所述音频信号的频率信息中的超过第一阈值的变化的模块;用于识别来自所述音频信号的振幅信息中的超过第二阈值的变化的模块;以及用于产生音符开始事件的模块,每个音符开始事件表示所述音频信号中的、所识别出的频率信息中的超过所述第一阈值的变化或所识别出的振幅信息中的超过所述第二阈值的变化中的至少一个的时间位置。

[0014] 从通过例子示出本发明原理的优选实施例的下列描述中,本发明的其他特征和优点应很明显。

[0015] 附图说明

[0016] 参考下列附图可实现对本发明的性质和优点的进一步的理解。在附图中,相似的部件或特征可具有相同的附图标记。进一步地,借助于在附图标记后继之以破折号和区分开相似部件的第二标记,可区分开相同类型的各种部件。如果在说明书中只使用第一附图标记,则此描述适用于具有相同的第一附图标记的相似部件中的任一个,而不考虑第二附图标记。

[0017] 图 1A 提供根据本发明的系统的高级简化方框图。

[0018] 图 1B 提供根据本发明的如同图 1 所示系统的系统的低级简化方框图。

[0019] 图 2 提供根据本发明实施例的用于将音频信号数据转换成总谱数据的示例性方法的流程图。

[0020] 图 3 提供根据本发明实施例的用于检测音高的示例性方法的流程图。

[0021] 图 4A 提供根据本发明实施例的用于产生音符开始事件的示例性方法的流程图。

[0022] 图 4B 提供根据本发明实施例的用于确定起声 (attack) 事件的示例性方法的流程图。

[0023] 图 5 提供根据本发明实施例的用在音符开始事件生成中的具有不同包络的音频信号的图示。

[0024] 图 6 提供根据本发明实施例的用于检测音符时值的示例性方法的流程图。

[0025] 图 7 提供根据本发明实施例的用在音符时值检测中的具有不同包络的音频信号的图示。

[0026] 图 8 提供根据本发明实施例的用于检测休止符 (rest) 的示例性方法的流程图。

[0027] 图 9 提供根据本发明实施例的用于检测拍速的示例性方法的流程图。

[0028] 图 10 提供根据本发明实施例的用于确定音符值的示例性方法的流程图。

[0029] 图 11 提供示出该示例性拍速检测方法的示例性数据的图表。

[0030] 图 12 提供示出图 11 所示的示例性拍速检测方法的额外的示例性数据。

[0031] 图 13 提供根据本发明实施例的用于检测音调的示例性方法的流程图。

[0032] 图 14A 和 14B 提供根据本发明实施例的用在音调检测中的两个示例性音调代价函数的图示。

[0033] 图 15 提供根据本发明实施例的用于确定音调音高标志的示例性方法的流程图。

[0034] 图 16 提供用于实现本发明的某些实施例的计算系统 1600 的结构图。

## 具体实施方式

[0035] 该描述仅提供示例性实施例,且不是用来限制本发明的范围、适用性或配置。更确切地,实施例的随后描述将给本领域技术人员提供用于实现本发明实施例的可行的描述。可在元素的功能和布置上进行各种变化,而不偏离本发明的实质和范围。

[0036] 因此,不同实施例可在适合时省略、替换或添加不同的程序或部件。例如,应认识到,在可选实施例中,可按不同于所描述的顺序执行方法,并可添加、省略或合并不同的步骤。此外,在不同的其他实施例中可合并关于某些实施例描述的特征。实施例的不同方面和元素可按类似的方式合并。

[0037] 还应认识到,下面的系统、方法和软件可单独地或共同为较大系统的部件,其中其他过程可优于其应用,或更改其应用。此外,在下面的实施例之前、之后或与其同时,可能需要多个步骤。

[0038] 图 1A 示出根据本发明构造的系统的高级简化方框图,该系统用于根据本发明自动和准确地提取来自音频信号的总谱数据。系统 100 在音频接收器单元 106 接收音频输入信号 104,并通过信号处理器单元 110、音符处理器单元 130 和总谱处理器单元 150 传递信号。总谱处理器单元 150 可接着产生总谱输出 170。

[0039] 根据本发明的一些实施例,系统 100 可接收音乐或表演作为音频输入信号 104,并产生表演的相应的音乐总谱表示 170。音频输入信号 104 可来自现场的表演或可包括已录制的表演的再现,并涉及乐器和人声部 (human voice) 两者。可为组成音频输入信号 104 的每一不同乐器和声部产生音乐总谱表示 170。音乐总谱表示 170 可提供例如音高、节奏、音色、力度强弱和 / 或任何其他有用的总谱信息。

[0040] 在一些实施例中,根据乐器和声部表演时 (例如,通过记录的区别) 的频率或通过区分开不同的音色,将单独地或组合地将乐器和声部与其他的辨别开。例如,在管弦乐队中,以不同的频率范围表演的个别乐师或乐师组 (例如,第一小提琴手或第二小提琴手,或小提琴手和大提琴手) 可被识别出并彼此区分开。类似地,麦克风或其他音频检测器的阵列可用于提高所接收的音频输入信号 104 的分辨率,增加包括在音频输入信号 104 中的音频声道或乐器的数量,或为音频输入信号 104 提供其他信息 (例如空间信息或深度)。

[0041] 在一个实施例中,音乐被麦克风或麦克风阵列 102 实时地接收,并转换成模拟电子音频输入信号 104,以由音频接收器单元 106 接收。在其他实施例中,音频输入信号 104 可包括数字数据,例如适合于再现的已录制的音乐文件。如果音频输入信号 104 是模拟信号,则它由音频接收器单元 106 变换成数字表示,以准备供信号处理器单元 110、音符处理器单元 130、总谱处理器单元 150 进行数字信号处理。因为输入信号被实时地接收,没有方法来预先确定音频输入信号 104 的完整长度。因此,音频输入信号 104 可以按预定的间隔被接收并储存 (例如,逝去时间的量、数字样本的数量、所使用的存储器的数量等),并可被相应地处理。在另一实施例中,已录制的声音片断由音频接收器 106 接收并被数字化,从而具有固定的持续时间。

[0042] 在一些实施例中,麦克风的阵列可用于检测同时演奏的多个乐器。将阵列中的各个麦克风放置成其比任何一个其他麦克风更接近于特定的乐器,因此由该乐器产生的频率的强度对该麦克风比对任何其他麦克风更高。在整个接收到的声音上合并四个检测器所提供

的信息,并使用所有麦克风所记录的信号,可导致音乐的数字抽象表示,在这种情况下这可模仿具有关于乐器的信息的录制的MIDI表示。信息的合并将包括涉及音高或音符的序列、频率的持续时间(节奏)、与基频相关的泛音系列(音色:乐器或特定的声部的类型)以及相对强度(力度强弱)的信息。可选地,单个麦克风可用于同时从多个乐器或其他源接收输出。

[0043] 在不同实施例中,从音频输入信号104提取的信息被处理,以自动产生音乐总谱表示170。传统软件包和库可用于从音乐总谱表示170产生活页乐谱。很多这样的工具接受以预定格式例如乐器数字接口(MIDI)等、以乐曲表示的形式的输入。因此,系统的一些实施例产生实质上与MIDI标准一致的音乐总谱表示170,以确保与这样的传统工具的相容性。一旦创建音乐总谱表示170,可能的应用就是大量的。在不同实施例中,总谱显示在设备显示器上、被打印出、被输入到音乐出版程序中、被储存或被与其他人共享(例如,对于合作的音乐项目)。

[0044] 应认识到,根据本发明,系统100的很多实现是可能的。在一些实施例中,系统100被实现为专用设备。该设备可包括一个或多个内部麦克风,其配置成检测声压,并将其转换成音频输入信号104,以被系统100使用。可选地,该设备可包括用于与外部麦克风、媒体设备、数据存储或其他音频源连接的一个或多个音频输入端口。在这些实施例中的某些中,该设备可为手持式或便携式设备。在其他实施例中,系统100可在多功能或通用设备中实现(例如,作为储存在计算机可读介质上的软件模块,用于由计算机执行)。在这些实施例中的某些中,音频源102可为声卡、外部麦克风或所储存的音频文件。音频输入信号104接着被产生并提供到系统100。

[0045] 系统100的其他实施例可被实现为简化或单声道形式,以作为从用户接收音频的听音设备来操作,这些用户对着一个麦克风演奏乐器或演唱某个曲调或旋律或其一部分。在单麦克风布置中,系统100相继地将来自一个麦克风的已录制的音乐转换成相应的音乐总谱。这可提供文本到语音软件的音乐等价物,所述文本到语音软件将口语单词和句子转换成计算机可读的文本。作为声音到音符转换,曲调或旋律将被记录为好像一个乐器正在演奏。

[0046] 应认识到,系统100的不同实现也可包括不同类型的接口和涉及与用户和其他系统的相容性的功能。例如,可为线路电平输入(例如,来自立体声系统或吉他放大器)、麦克风输入、网络输入(例如,来自互联网)或其他数字音频部件提供输入端口。类似地,可为扬声器、音频部件、计算机和网络等的输出提供输出端口。进一步地,在一些实现中,系统100可提供用户输入(例如,物理或虚拟键盘、滑块、旋钮、开关等)和/或用户输出(例如,显示器、扬声器等)。例如,可提供接口能力,以允许用户听录制品或通过系统100从录制品提取的数据。

[0047] 在图1B中提供了系统100的一个实施例的低级结构图。一个或多个音频源102可用于产生音频输入信号。音频源102可为能够向音频接收器106提供音频输入信号104的任何源。在一些实施例中,一个或多个麦克风、转换器、和/或其他传感器用作音频源102。麦克风可将来自现场表演(或已录制的表演的再现)的压力或电磁波转换成电信号,以用作音频输入信号104。例如,在现场的音频表演中,麦克风可用于检测和转换来自歌手的音频,而电磁“拾音器(pick-up)”可用于检测和转换来自吉他和贝司的音频。在其他实施例

中,音频源 102 可包括配置成提供音频输入信号 104 或音频文件的模拟或数字设备,从该音频文件可读取音频输入信号 104。例如,数字化音频文件能够以音频格式储存在存储介质上,并作为音频输入信号 104 由存储介质提供给音频接收器 106。

[0048] 应认识到,根据音频源 102,音频输入信号 104 可具有不同的特征。音频输入信号 104 可为单音调或复调的,可包括音频数据的多个声道,可包括来自很多类型的乐器的音频,并可包括特定文件格式,等等。类似地,应认识到,音频接收器 106 可为能够接收音频输入信号 104 的任何接收器。进一步地,音频接收器 106 可包括一个或多个端口、解码器、或与音频源 102 连接、或接收或解释音频输入信号 104 所必需的其他部件。

[0049] 音频接收器 106 可提供额外的功能。在一个实施例中,音频接收器 106 将模拟音频输入信号 104 转换成数字音频输入信号 104。在另一实施例中,音频接收器 106 配置成将音频输入信号 104 下变频到较低的采样率,以减小系统 100 的计算负担。在一个实施例中,音频输入信号 104 被下采样到大约 8-9kHz。这可提供音频输入信号 104 的较高的频率分辨率,并可减小对系统 100 的设计的某些限制(例如,滤波器规范)。

[0050] 在又一实施例中,音频接收器 106 包括阈值检测部件,其配置成在检测到音频电平超过某个阈值时开始接收音频输入信号 104(例如,开始录制)。例如,阈值检测部件可在规定的一段时间内分析音频,以检测音频输入信号 104 的振幅是否在某个预定量的时间内保持在预定阈值之上。阈值检测部件可进一步配置成当音频输入信号 104 的振幅在预定量的时间内落在预定阈值之下时停止接收音频输入信号 104(例如,停止录制)。在又一实施例中,阈值检测部件可用于为系统 100 产生标志,其表示音频输入信号 104 的振幅在一定量的时间内超过阈值或落在阈值之下的条件,而不是实际上开始或终止音频输入信号 104 的接收。

#### [0051] 信号和音符处理

[0052] 根据图 1B,音频接收器 106 将音频输入信号 104 传递到包括振幅提取单元 112 和频率提取单元 114 的信号处理器单元 110。振幅提取单元 112 配置成从音频输入信号 104 提取与振幅有关的信息。频率提取单元 114 配置成从音频输入信号 104 提取与频率有关的信息。

[0053] 在一个实施例中,频率提取单元 114 使用变换算法将信号从时域变换到频域。例如,当在时域中时,音频输入信号 104 可被表示为振幅随时间而变化。然而,在应用快速傅立叶变换(FFT)算法之后,同一音频输入信号 104 可被表示为其每个频率分量的振幅的曲线(例如,在频率范围内的每个频带的相对强度或组成,如泛音系列,信号将在其上被处理)。对于处理效率,可能希望将算法限制到某个频率范围。例如,频率范围可只覆盖可听声谱(例如,大约 20Hz 到 20kHz)。

[0054] 在不同实施例中,信号处理器单元 110 可用其他方式提取与频率有关的信息。例如,很多变换算法以固定宽度的线性频率“存储段(bucket)”输出信号。这可限制变换的可能频率分辨率或有效性,特别是考虑到音频信号在本质上内在地可以是对数的(而不是线性的)。用于从音频输入信号 104 提取与频率有关的信息的很多算法在本领域中是已知的。

[0055] 振幅提取单元 112 所提取的与振幅有关的信息和频率提取单元 114 所提取的与频率有关的信息可接着由音符处理单元 130 的不同部件使用。在一些实施例中,音符处理单

元 130 包括音符开始检测器单元 132、音符时值检测器单元 134、音高检测器单元 136、休止符检测器单元 144、包络检测器单元 138、音色检测器单元 140 和音符力度强弱检测器单元 142 中的全部或一些。

[0056] 音符开始检测器单元 132 配置成检测音符的开始。音符的开始（或开始）一般在音乐中表现为音高的变化（例如圆滑奏）、振幅的变化（例如包络的附加部分）或音高和振幅的变化的一些组合。因此，音符开始检测器单元 132 可配置成每当有频率（或音高）和 / 或振幅的某种类型的变化时，则产生音符开始事件，如下面关于图 4-5 更详细描述。

[0057] 音乐音符也可以其时值（例如，音符以秒或样本的数量为单位持续的的时间的量）为特征。在一些实施例中，音符处理单元 130 包括音符时值检测器单元 134，其配置成检测音符开始事件所标记的音符的时值。下面关于图 6 和 7 更详细地讨论音符时值的检测。

[0058] 值得注意的是，音乐的某些特征是心理声学的，而不纯粹地是信号的物理属性。例如，频率是信号的物理特性（例如，表示正弦波每秒所传播的周期数），但音高是更复杂的心理声学现象。一个原因是乐器所演奏的单个音高的音符通常由多个频率组成，每个频率振幅不同，被称为音色。大脑可感知那些频率之一（例如，一般是基频）作为“音高”，而感测其他频率仅是向给符添加“和声色彩”。在一些情况下，听者所感受到的音符的音高可能为信号最缺乏或完全缺乏的频率。

[0059] 在一些实施例中，音符处理单元 130 包括音高检测器单元 136，其配置成检测由音符开始事件标记的音符的音高。在其他实施例中，音高检测器单元 136 配置成跟踪音频输入信号 104 的音高，而不是（或除了）跟踪个别音符的音高。应认识到，音高检测器单元 136 在一些情况下可由音符开始检测器单元 132 使用，以确定音频输入信号 104 的超过阈值的音高的变化。

[0060] 音高检测器单元 136 的某些实施例进一步处理音高，以与最后的音乐总谱表示 170 更相容。关于图 3 更充分地描述音高检测的实施例。

[0061] 音符处理单元 130 的一些实施例包括配置成检测音频输入信号 104 内休止符的存在的休止符检测器单元 144。休止符检测器单元 144 的一个实施例使用振幅提取单元 112 所提取的与振幅有关的信息和音高检测器单元 136 所得到的置信度信息。例如，与振幅有关的信息可揭示音频输入信号 104 的振幅在某个时间窗内相对低（例如，在噪声基底处或附近）。在相同的时间窗内，音高检测器单元 136 可确定任何特定音高的存在的非常低的置信度。使用这个和其他信息，休止符检测器单元 144 检测休止符的存在以及休止符可能开始的时间位置。关于图 9 和 10 进一步描述检测休止符的实施例。

[0062] 在一些实施例中，音符处理单元 130 包括音色检测器单元 140。振幅提取单元 112 所提取的与振幅有关的信息和频率提取单元 114 所提取的与频率有关的信息可由音色检测器单元 140 用来对音频输入信号 104 的一部分检测音色信息。音色信息可揭示音频信号 104 的该部分的和声成分。在一些实施例中，音色检测器单元 140 可检测关于在音符开始事件处开始的特定音符的音色信息。

[0063] 在音色检测器单元 140 的一个实施例中，与振幅有关的信息和与频率有关的信息与高斯滤波器卷积，以产生滤波的频谱。滤波的频谱可接着用于产生在音高检测器单元 136 所检测的音高周围的包络。该包络可相应于在该音高处的音符的音色。

[0064] 在一些实施例中，音符处理单元 130 包括包络检测器单元 138。振幅提取单元 112

所提取的与振幅有关的信息可由包络检测器单元 138 使用来对音频输入信号 104 的一部分检测包络信息。例如,击钢琴上的键可使音槌击打一组弦,导致具有大起声振幅的音频信号。该振幅经历快速衰减,直到它在弦共振的地方维持在稍微稳态的振幅(当然,当弦中的能量耗尽时,振幅可在包络的该部分上缓慢地减小)。最后,当钢琴键被释放时,制音器落在弦上,使振幅快速下降到零。这种类型的包络一般称为 ADSR(起声(attack)、衰减(decay)、维持(sustain)、释放(release))包络。包络检测器单元 138 可配置成检测 ADSR 包络的一些或所有部分,或任何其他类型的有用包络信息。

[0065] 在不同实施例中,音符处理单元 130 也包括音符力度强弱检测器单元 142。在某些实施例中,音符力度强弱检测器单元 142 为在某些音符开始事件处开始的特定音符提供与包络检测器单元 138 的功能类似的功能。在其他实施例中,音符力度强弱检测器单元 142 配置成检测相对于被包络检测器单元 138 检测到的包络图形是异常的或符合某种预定的模式(pattern)的音符包络。例如,断音音符可以以急剧的起声和其 ADSR 包络的短持续部分为特征。在另一例子中,重音音符可以明显大于周围音符的起声振幅的起声振幅为特征。

[0066] 应认识到,音符力度强弱检测器单元 142 和其他音符处理单元可用于识别音符的多个其他属性,这些属性可被描述为音乐总谱表示 170 的部分。例如,音符可被标记为圆滑奏、重音、断音、装饰音符等。根据本发明可提取很多其他音符特征。

#### [0067] 总谱处理

[0068] 与多个音符或音符开始事件(包括休止符)有关的信息可用于产生其他信息。根据图 1B 的实施例,音符处理单元 130 的不同部件可与总谱处理单元 150 的不同部件操作性地通信。总谱处理单元 150 可包括拍速检测单元 152、节拍检测单元 154、音调检测单元 156、乐器识别单元 158、声道检测单元 162 和全局力度强弱检测单元 164 中的全部或一些。

[0069] 在一些实施例中,总谱处理单元 150 包括拍速检测单元 152,其配置成在一时间窗内检测音频输入信号 104 的拍速。一般,一首音乐的拍速(例如,音乐似乎在心理声学上传递的速度)可部分地受到音符和休止符的存在和时值的影响。因此,拍速检测单元 152 的某些实施例使用来自音符开始检测器单元 132、音符时值检测器单元 134 和休止符检测器单元 144 的信息来确定拍速。拍速检测单元 152 的其他实施例进一步使用确定的拍速来给音符和休止符分配音符值(例如,四分音符、八分音符等)。相关于图 11-15 进一步详细地讨论拍速检测单元 152 的示例性操作。

[0070] 节拍指示在音乐的每个小节中有多少拍,以及哪个音符值被认为是单拍。例如,4/4 的拍子表示每个小节有四拍(分子),以及单拍由四分音符(分母)表示。由于这个原因,节拍可帮助确定音符和小节线位置,以及提供有用的音乐总谱表示 170 可能需要的其他信息。在一些实施例中,总谱处理单元 150 包括配置成检测音频输入信号 104 的节拍的节拍检测单元 154。

[0071] 在一些实施例中,从拍速检测单元 152 所提取的音符值和拍速信息以及从其他信息(例如,音符力度强弱检测器单元 142 所提取的音符力度强弱信息)推断简单的节拍。然而,通常,确定节拍是涉及复杂的模式识别的复杂任务。

[0072] 例如,假定音符值的下列序列从音频输入信号 104 被提取:四分音符、四分音符、八分音符、八分音符、八分音符、八分音符。此简单的序列可被表示为 4/4 的一个小节、2/4 的两个小节、1/4 的四个小节、8/8 的一个小节或很多其他小节。假定在第一个四分音符和



第一个八分音符上有重音（例如，增加的起声振幅），这可使序列更可能为 2/4 的两个小节、4/8 的两个小节或 4/4 的一个小节。进一步地，假定 4/8 是非常罕见的拍子，则可足以消除此猜测。更进一步地，若知道音频输入信号 104 的类型是民谣，则最可能的候选节拍更可能是 4/4。

[0073] 上面的例子示出涉及甚至非常简单的音符值序列的复杂性。很多音符序列复杂得多，涉及不同值的很多音符、跨越多个小节的音符、附点音符和装饰音符、切分音以及在解释拍子方面的其他困难。由于这个原因，传统计算算法可能在准确地确定拍子上有困难。因此，节拍检测单元 154 的不同实施例使用被训练来检测那些复杂模式的人工神经网络 (ANN) 0160。通过给 ANN 0160 提供不同拍子的很多样本和使用每个样本改进的代价函数可训练 ANN 0160。在一些实施例中，使用学习范式来训练 ANN 0160。学习范式可包括例如有监督的学习、没有监督的学习或强化学习算法。

[0074] 应认识到，通过使用拍速和节拍信息中的任一个或两个可产生很多有用类型的信息，以被音乐总谱表示 170 使用。例如，信息可允许确定：在哪里将音符分为小节（例如，作为八分音符组），而不是使用符尾单独地指定音符；何时将一音符分在两个小节中，并将其用连接线连起来；或何时将音符组指定为三连音符（或更高阶的组）、装饰音符、颤音或波音、滑音等。

[0075] 在产生音乐总谱表示 170 中可能有用的另一组信息涉及音频输入信号 104 的一部分的音调。音调信息可包括例如被识别的根音高和相关调式。例如，“A 小调”表示音调的根音高是“A”，而调式是小调。每个音调以调号为特征，调号标识“在音调中”（例如，与音调关联的自然音阶的部分）和“在音调外”（例如，在音调的范式中的临时符号）的音符。“A 小调”例如不包含升号 (sharp) 或降号 (flat)，而“D 大调”包含两个升号且不包含降号。

[0076] 在一些实施例中，总谱处理单元 150 包括配置成检测音频输入信号 104 的音调的音调检测单元 156。音调检测单元 156 的一些实施例基于比较音高序列与一组代价函数来确定音调。代价函数可例如试图在特定的时间窗内最小化一首音乐中的临时符号的数量。在其他实施例中，音调检测单元 156 可使用人工神经网络来进行或改进复杂的音调确定。在又一些其他实施例中，可对照代价函数来评估音调变化的序列，以改进音调确定。在另一些其他实施例中，音调检测单元 156 所得到的音调信息可用于将音符（或音符开始事件）赋予特定的音调音高标志。例如，F 大调中的“B”可被指定为“B- 本位音”。当然，音调信息可用于产生调号或音乐总谱表示的其他信息。在一些实施例中，音调信息可进一步用于产生和弦或其他和声信息。例如，以弦线乐谱的格式可产生吉他和弦，或可提供爵士和弦。关于图 13-15 进一步详细描述音调检测单元 156 的示例性操作。

[0077] 在其他实施例中，总谱处理单元 150 还包括配置成识别音频输入信号 104 的正演奏的乐器的乐器识别单元 158。乐器常常被认为具有特定的音色。然而，根据被演奏的音符或音符被演奏的方式，在单个乐器上可能有音色的差异。例如，根据例如在其构造中使用的材料、表演者的触摸、被演奏的音符（例如，在开弦上演奏的音符具有与在用手指弹奏的弦上演奏的同一音符不同的音色，以及在小提琴音区下部的音符具有与上部音区中的音符不同的音色）、音符是用弓拉的还是弹拨的等，每个小提琴的音色不同。然而，在小提琴音符之间仍然可能有足够的类似，以将它们识别为小提琴，而不是其他乐器。

[0078] 乐器识别单元 158 的实施例配置成比较单个或多个音符的特征，以确定明显由音

频输入信号 104 的乐器表演的音高的音域、由乐器在那些音高的每个音高处产生的音色、和 / 或在乐器上演奏的音符的振幅包络。在一个实施例中,音色差异被用于通过将乐器样本的一般音色符号与来自音频输入信号 104 的检测到的音色进行比较,来检测不同的乐器。例如,即使以相同音量演奏相同时值的相同音符时,萨克斯和钢琴由于其不同的音色也可能发出非常不同的声音。当然,如上所述,只基于音色的识别的准确性可能有限。

[0079] 在另一实施例中,音高音域 (pitch range) 用于检测不同的乐器。例如,大提琴一般可演奏范围从中央 C 音之下的大约两个八度到中央 C 音之上的大约一个八度的音符。然而,小提琴一般可演奏范围从恰好在中央 C 音之下到中央 C 音之上的大约四个八度的音符。因此,即使小提琴和大提琴可具有相似的音色(它们都是拉奏弦乐器),其音高音域可充分不同以用于识别。当然,考虑到音域在某种程度上确实交叠,可能会有误差。进一步地,其他乐器(例如钢琴)具有可与很多乐器交叠的较大音域。

[0080] 在又一实施例中,包络检测用于识别不同的乐器。例如,在槌打乐器(例如钢琴)上演奏的音符可能发声不同于在木管乐器(例如长笛)、簧乐器(例如双簧管)、铜管乐器(例如小号)或弦乐器(例如小提琴)上演奏的同一音符。然而,每种乐器能够产生很多不同类型的包络,取决于音符如何被演奏。例如,小提琴可被弹拨或拉奏,或音符可被连奏或断奏。

[0081] 至少由于上面提到的困难,准确的乐器识别可能需要检测复杂的模式,包括可能在多个音符上的音频输入信号 104 的多个特征。因此,乐器识别单元 158 的一些实施例利用被训练来检测这些复杂模式的组合的人工神经网络。

[0082] 总谱处理单元 150 的一些实施例包括配置成识别来自音频输入信号 104 内的音频声道的声道检测单元 162。在一些情况下,音频输入信号 104 可以为已经被声道分离的格式。例如,在某些数字音频磁带 (DAT) 上的音频可被储存在 8 个分离的数字音频声道。在这些情况下,声道检测单元 162 可配置成仅仅识别单独的音频声道。

[0083] 然而在其他情况下,多个声道可储存在单个音频输入信号 104 中,且需要通过从音频输入信号提取某些数据来进行识别。因此,声道检测单元 162 的一些实施例配置成,使用从音频输入文件 104 提取的信息来识别分开的音频声道。例如,表演可包括同时演奏的 5 个乐器(例如爵士五重奏)。可能希望将那些分开的乐器识别为分开的声道,以能够在音乐总谱表示 170 中准确地表示表演。

[0084] 声道检测可用多种不同的方法完成。在一个实施例中,声道检测单元 162 使用音高检测来确定是否不同的音符序列表现为被限制到某些音高音域。在另一实施例中,声道检测单元 162 使用来自乐器识别单元 158 的乐器识别信息来确定不同的声道。

[0085] 很多总谱也包含关于音乐或演奏的全局力度强弱的信息。全局力度强弱指跨越多个音符的力度强弱,而与上述音符力度强弱不同。例如,整个乐段 (piece) 或乐段的部分可被标记为强的(高声)或弱的(轻柔)。在另一例子中,音符的序列可以声音渐强逐渐增大。为了产生这种类型的信息,总谱处理单元 150 的一些实施例包括全局力度强弱检测单元 164。全局力度强弱检测单元 164 的实施例使用振幅信息,在一些情况下包括音符力度强弱信息和 / 或包络信息,以检测全局力度强弱。

[0086] 在某些实施例中,从音频输入信号 104 预先确定或自适应地产生阈值,以有助于力度强弱确定。例如,摇滚表演的平均音量可被认为是强的。超过该平均值某个数量(例

如, 阈值、标准偏差等) 的振幅可被认为是极强的, 而落在该平均值之下某个数量的振幅可被认为是弱的。

[0087] 某些实施例可进一步考虑力度强弱变化出现的时值。例如, 以两分钟的平静音符开始并突然转换到两分钟的较大声的音符段的乐段可被认为具有跟随有强音的部分的弱音的部分。另一方面, 在几个音符的过程中增加、对几个更多的音符保持在该较高的音量并接着返回到原始振幅的平静乐段可被认为具有跟随有渐弱部分的渐强部分。

[0088] 可产生上面描述的所有不同类型的信息以及任何其他有用的信息, 以用作音乐总谱表示 170。该音乐总谱表示 170 可被保存或输出。在某些实施例中, 音乐总谱表示 170 输出到总谱生成软件, 该软件可将各种类型的信息转录成总谱格式。该总谱格式可配置成供观看、打印、电传输等。

[0089] 应认识到, 上面描述的各种单元和部件可用各种方法实现, 而不偏离本发明。例如, 某些单元可为其他单元的部件, 或可实现为另一单元的额外功能。进一步地, 根据本发明, 可用很多方法连接单元, 且数据可在其间以很多方式流动。因此, 图 1B 应被理解为说明性的, 且不应被解释为限制本发明的范围。

#### [0090] 用于音频处理的方法

[0091] 图 2 提供了根据本发明实施例的用于将音频信号数据转换成总谱数据的示例性方法的流程图。方法 200 通过接收音频信号在块 202 开始。在一些实施例中, 可预处理音频信号。例如, 音频信号可从模拟转换到数字、下变频到较低的采样率、为与某些编码器或解码器的相容而被转换代码、解析成单音调音频声道、或任何其他有用的预处理。

[0092] 在块 204, 可从音频信号提取频率信息, 并可识别频率中的特定变化。在块 206, 可从音频信号提取振幅信息, 并可识别振幅中的变化。

[0093] 在一些实施例中, 在块 208 中从频率信息得到音高信息, 该频率信息在块 204 中从音频输入信号提取。关于图 3 更充分地描述了在块 208 的音高检测的示例性实施例。进一步地, 在一些实施例中, 在块 210, 与频率和振幅有关的所提取和识别的信息用于产生音符开始事件。关于图 4-5 更充分地描述了在块 210 的音符开始事件生成的示例性实施例。

[0094] 在方法 200 的一些实施例中, 在块 204 中提取的频率信息、在块 206 中提取的振幅信息以及在块 210 中产生的音符开始事件用于提取并处理来自音频信号的其他信息。在某些实施例中, 上述信息用于在块 220 确定音符时值、在块 230 确定休止符、在块 240 确定时间窗上的拍速、在块 250 确定窗上的音调、以及在块 260 确定乐器。在其他实施例中, 在块 220 确定的音符时值、在块 230 确定的休止符以及在块 240 确定的拍速用于在块 245 确定音符值; 在块 250 确定的音调用于在块 255 确定音调音高标志; 以及在块 260 确定的乐器用于在块 270 确定声道。在不同实施例中, 块 220-270 的输出配置成用于在块 280 产生音乐总谱表示数据。关于图 6-15 更详细描述了块 220-255 的示例性方法。

#### [0095] 音高检测

[0096] 图 3 提供了根据本发明实施例的用于检测音高的示例性方法的流程图。音高的人类感知是心理声学现象。因此, 方法 208 的一些实施例通过使用心理声学滤波器组预滤波音频输入信号而在块 302 开始。在块 302 的预滤波可涉及例如模拟人耳的听觉范围的加权音阶。这种加权音阶对本领域技术人员是已知的。

[0097] 方法 208 接着可通过将音频输入信号 104 分成预定的音程在块 304 继续。这些音

程可以基于音符开始事件、信号的采样频率或任何其他有用的音程。根据音程类型,方法 208 的实施例可例如配置成检测音符开始事件所标记的音符的音高或跟踪音频输入信号中的音高变化。

[0098] 对于每个音程,方法 208 可在块 306 检测基频。基频可被指定为音程的(或音符的)“音高”。基频常常是最低有效频率和具有最大强度的频率,但不是总是这样。

[0099] 方法 208 可进一步处理音高以与最后的音乐总谱表示更相容。例如,音乐总谱表示可能需要定义明确和有限的音高组,其由组成总谱的音符表示。因此,方法 208 的实施例可将频谱分成与特定的音乐音符相关的盒(bin)。在一个实施例中,方法 208 计算每个盒中的能量,并将具有最低有效能量的盒识别为基本音高频率。在另一实施例中,方法 208 根据每个盒中的能量计算音频输入信号的泛音系列,并使用该泛音系列来确定基本音高频率。

[0100] 在示例性实施例中,方法 208 使用具有一组均匀交叠的两个八度宽的滤波器的滤波器组。每个滤波器组应用于音频输入信号的一部分。每个滤波器组的输出被分析,以确定音频输入信号的被滤波部分是否是充分正弦形的,以本质上包含单一频率。以这种方式,方法 208 能够在某个时间间隔内提取音频输入信号的基频,作为在该音程期间的信号的音高。在某些实施例中,方法 208 可配置成在一音程内得到音频输入信号的基频,即使在该信号中没有基频的情况下(例如,通过使用在该时间窗期间存在于音频输入信号中的频率的泛音系列中的几何关系)也是如此。

[0101] 在一些实施例中,方法 208 在块 308 使用一系列滤波器组输出来产生一组音频样本。每个音频样本可具有相关的数据记录,包括例如关于所评估的频率、置信度值、时间戳、时值和钢琴音调指标的信息。应认识到,用于从音频输入信号提取该数据记录信息的很多方法在本领域中是已知的。在 Lawrence Saul、Daniel Lee、Charles Isbell 和 Yaun LeCun 的“Real time voice processing with audiovisual feedback: toward autonomous agents with perfect pitch,” Advances in Neural Information Processing Systems (NIPS) 15, pp. 1205-1212 (2002) 中详述了一种示例性方法,其这里为了所有目的通过引用而被并入。音频样本的数据记录信息可被缓存和储存以确定听者将听到什么音高。

[0102] 方法 208 的一些实施例通过确定音高变化出现在哪里来在块 310 继续。例如,如果音高分成音乐盒(例如音阶音),可能希望确定音频信号的音高在何处从一个盒跨到下一盒。否则,颤音、震音和其他音乐效果可能被错误地识别为音高变化。识别音高变化的开始在确定音符开始事件中也是有用的,如下所述。

#### [0103] 音符开始检测

[0104] 音乐乐曲的很多元素至少部分地以音符的开始为特征。例如在总谱上,可能必须知道音符在哪里开始,以确定音符在小节中的正确的时间位置、乐曲的拍速和节拍、以及其他重要的信息。一些抒情的音乐表演包括音符变化,其涉及音符在哪里开始的主观确定(例如,由于从一个音符到另一个的缓慢圆滑奏)。然而,总谱生成可能需要更客观地确定音符在哪里开始和结束。这些音符开始在这里称为音符开始事件。

[0105] 图 4A 提供了根据本发明实施例的用于产生音符开始事件的示例性方法的流程图。方法 210 通过识别音高变化事件在块 410 开始。在一些实施例中,根据从(例如,如在图 2 的块 204 中的)音频信号提取的频率信息 402 中的超过第一阈值 404 的变化,在块 410 确定音高变化事件。在方法 210 的一些实施例中,使用参考图 2 的块 208 描述的方法来识

别音高变化事件。

[0106] 通过在块 410 识别音高变化事件,每当在音高中有充分的变化时,方法 210 可在块 450 检测音符开始事件。以这种方式,即使对于在振幅中没有可检测到的变化的从一个音高到另一个的缓慢圆滑奏,也将在块 450 产生音符开始事件。然而只使用音高检测将不能检测重复的音高。如果表演者多次演奏一行中的相同音高,则在音高中没有变化,而导致未在块 410 用信号通知音高变化事件,以及在块 450 不产生音符开始事件。

[0107] 因此,方法 210 的实施例也在块 420 识别起声事件。在一些实施例中,根据从(例如,如在图 2 的块 206 中的)音频信号提取的振幅信息 406 中的超过第二阈值 408 的变化在块 420 确定起声事件。起声事件可为音频信号的振幅中的特征性变化,以用信号通知音符的开始。通过在块 420 识别起声事件,每当在振幅中有特征性变化时,方法 210 可在块 450 检测音符开始事件。以这种方式,即使重复的音高也将在块 450 产生音符开始事件。

[0108] 应认识到,用于检测起声事件的很多方法是可能的。图 4B 提供了根据本发明实施例的用于确定起声事件的示例性方法的流程图。方法 420 在块 422 通过使用从音频信号提取的振幅信息 406 来产生第一包络信号而开始。第一包络信号可表示跟踪音频信号的振幅中的包络电平变化的“快速包络”。

[0109] 在一些实施例中,通过首先整流并滤波振幅信息 406 来在块 422 产生第一包络信号。在一个实施例中,采取信号振幅的绝对值,该绝对值接着使用全波整流器被整流以产生音频信号的整流版本。接着可通过使用低通滤波器滤波被整流的信号来产生第一包络信号。这可产生实质上持有被整流的音频信号的整体形式的第一包络信号。

[0110] 第二包络信号可在块 424 产生。第二包络信号可表示近似于音频信号的包络的平均功率的“缓慢包络”。在一些实施例中,通过连续地或在预定时间间隔内(例如通过对信号积分)计算第一包络信号的平均功率可在块 424 产生第二包络信号。在某些实施例中,第二阈值 408 可在给定的时间位置从第二包络信号的值得到。

[0111] 在块 426,产生控制信号。该控制信号可表示第一包络信号中更明显的方向变化。在一个实施例中,控制信号通过下列操作在块 426 产生:(1)在第一时间位置找到第一包络信号的振幅;(2)在该振幅处继续一直到第二时间位置(例如,第一和第二时间位置被间隔开预定量的时间);以及(3)将第二时间位置设置为新的时间位置并重复此过程(即,移到第二时间位置处的新振幅并在预定数量的时间内保持在那里)。

[0112] 方法 420 接着在块 428 将控制信号变得大于(例如,在正方向上跨过)第二包络信号的任何位置识别为起声事件。以这种方式,可以只在包络中的明显变化出现的地方识别起声事件。该方法 420 的示例性图示在图 5 中示出。

[0113] 图 5 提供了根据本发明实施例的用在音符开始事件产生中的具有不同包络的音频信号的图示。例证性曲线 500 描绘音频输入信号 502、第一包络信号 504、第二包络信号 506 和控制信号 508 的振幅与时间的关系曲线。该曲线也示出起声事件位置 510,在该位置 510,控制信号 508 的振幅变得大于第二包络信号 506 的振幅。

#### [0114] 音符时值检测

[0115] 一旦通过产生音符开始事件识别出音符的开始,确定音符在何处结束(或音符的时值)就可能是有用的。图 6 提供了根据本发明实施例的用于检测音符时值的示例性方法的流程图。方法 220 通过在块 602 识别第一音符开始位置而开始。在一些实施例中,通过

产生（或识别）音符开始事件在块 602 识别第一音符开始位置，如关于图 4-5 更充分描述的。

[0116] 在一些实施例中，方法 220 通过在块 610 识别第二音符开始位置而继续。可用与在块 602 识别第一音符开始位置相同或不同的方法来在块 610 识别该第二音符开始位置。在块 612，通过确定第一音符开始位置到第二音符开始位置之间的时间间隔来计算与第一音符开始位置相关的音符的时值。在块 612 中的该确定可产生音符的时值，作为从一个音符的开始到下一音符的开始的逝去时间。

[0117] 然而在一些情况下，一音符可在下一音符开始之前的某个时间结束。例如，一音符可跟随有休止符，或音符能够以断音方式被表演。在这些情况下，在块 612 中的该确定可产生超过音符的实际时值的音符时值。值得注意的是，可通过检测音符结束位置以很多方法校正这个可能的限制。

[0118] 方法 220 的一些实施例在块 620 识别音符结束位置。在块 622，接着通过确定第一音符开始位置和音符结束位置之间的时间间隔可计算与第一音符开始位置相关的音符的时值。在块 622 中的该确定可产生音符的时值，作为从一个音符的开始到该音符的结束的逝去时间。一旦在块 612 或在块 622 确定了音符时值，音符时值就可在块 630 被分配给在第一时间位置开始的音符（或音符开始事件）。

[0119] 应认识到，根据本发明用于在块 620 识别音符结束位置的很多方法是可能的。在一个实施例中，通过确定在音符之间是否存在任何休止符并从音符时值减去休止符的时值（下面讨论休止符的检测和休止符时值）来在块 620 检测音符结束位置。在另一实施例中，分析音符的包络以确定音符是否以这样的方式被演奏，以便改变其时值（例如，以断音方式）。

[0120] 在块 620 的又一实施例中，以类似于在图 4B 的方法 420 中检测音符开始位置的方式检测音符结束位置。使用从音频输入信号提取的振幅信息，第一包络信号、第二包络信号和控制信号都可产生。通过识别控制信号的振幅变得小于第二包络信号的振幅的位置可确定音符结束位置。

[0121] 值得注意的是，在复调音乐中，可能有音符交叠的情况。因此，可能有第一音符的结束出现在第二音符的开始之后但在第二音符的结束之前的状况。因此，简单地在音符开始之后检测第一音符结束可能不能为该音符产生适当的结束位置。因此，可能必须提取单音调声道（如下所述）以更准确地识别音符时值。

[0122] 图 7 提供了根据本发明实施例的用在音符时值检测中的具有不同包络的音频信号的图示。该例证性曲线 700 描绘音频输入信号 502、第一包络信号 504、第二包络信号 506 和控制信号 508 的振幅与时间的关系曲线。该曲线也示出音符开始位置 710 和音符结束位置 720，在位置 710，控制信号 508 的振幅变得大于第二包络信号 506 的振幅，在位置 720，控制信号 508 的振幅变得小于第二包络信号 506 的振幅。

[0123] 曲线 700 进一步示出音符时值检测的两个实施例。在一个实施例中，通过找到第一音符开始位置 710-1 和第二音符开始位置 710-2 之间的逝去时间来确定第一音符时值 730-1。在另一实施例中，通过找到第一音符开始位置 710-1 和第一音符结束位置 720-1 之间的逝去时间来确定第二音符时值 740-1。

[0124] 休止符检测

[0125] 图 8 提供了根据本发明实施例的用于检测休止符的示例性方法的流程图。方法 230 通过在块 802 识别输入音频信号中的低振幅状态而开始。应认识到,根据本发明用于识别低振幅状态的很多方法是可能的。在一个实施例中,将噪声阈值电平设置在高于输入音频信号的噪声基底的某个振幅处。低振幅状态可接着被识别为输入音频信号的一区域,在此区域期间,该信号的振幅在某个预定量的时间内保持低于该噪声阈值。

[0126] 在块 804,分析存在低振幅状态的区域的音高置信度 (pitchconfidence)。音高置信度可识别音高 (例如,作为预期音符的部分) 存在于该区域中的可能性。应认识到,可用很多方法确定音高置信度,例如参考上面的音高检测描述的。

[0127] 在信号的低振幅区域中音高置信度低于某个音高置信度阈值的地方,很有可能不存在任何音符。在某些实施例中,在块 806,确定不存在音符的区域包括休止符。当然,如上所述,其他音乐条件可导致休止符的出现 (例如断音音符)。因此,在一些实施例中,其他信息 (例如,包络信息、乐器识别等) 可用于改进休止符是否存在的判断。

#### [0128] 拍速检测

[0129] 一旦知道音符和休止符的位置,就可能希望确定拍速。拍速使拍的适应性音乐概念与时间的标准物理概念匹配,本质上提供音乐乐曲的速度的度量 (例如,音乐应被演奏得多快)。拍速常常以每分钟拍子的数量表示,其中拍子由某个音符值表示。例如,音乐总谱可将单拍表示为四分音符,且拍速可为 84 拍每分钟 (bpm)。在本例中,以指定的拍速表演乐曲意味着以一速度演奏乐曲,在该速度,相当于每分钟表演音乐的 84 个四分音符。

[0130] 图 9 提供了根据本发明实施例的用于检测拍速的示例性方法的流程图。方法 240 通过在块 902 确定一组参考拍速而开始。在一个实施例中,可使用标准节拍器拍速。例如,一般节拍器可配置成保持范围从 40bpm 到 208bpm 的 4bpm 音程 (即,40bpm,44bpm,48bpm, … 208bpm) 的拍速的时间。在其他实施例中,可使用其他值和价值之间的音程。例如,这组参考拍速可包括范围从 10bpm 到 300bpm 的 1/4bpm 音程 (即,10bpm,10.25bpm,10.5bpm, … 300bpm) 所有拍速。

[0131] 方法 240 可接着确定每个参考拍速的参考音符时值。参考音符时值可表示某个音符值以给定的参考拍速持续多长时间。在一些实施例中,可以时间 (例如,秒) 测量参考音符时值,而在其他实施例中,可以样本的数量测量参考音符时值。例如,假定四分音符表示单拍,84bpm 的四分音符将持续大约 0.7143 秒 (即,每分钟 60 秒除以每分钟 84 拍)。类似地,假定每秒 44,100 个样本的采样率,在 84bpm 的四分音符将持续 31,500 个样本 (即,每秒 44,100 个样本乘以每分钟 60 秒除以每分钟 84 拍)。在某些实施例中,可在每个参考拍速处评估多个的音符值,以产生这组参考音符时值。例如,十六分音符、八分音符、四分音符和二分音符都可被评估。以这种方式,可为每个参考拍速产生理想化的音符值。

[0132] 在方法 240 的一些实施例中,可在块 906 确定拍速提取窗。拍速提取窗可为跨越音频输入信号的某个连续部分的预先确定的或自适应的时间窗。优选地,拍速提取窗宽到足以覆盖大量音符开始事件。因此,块 906 的某些实施例调节拍速提取窗的宽度,以覆盖预定数量的音符开始事件。

[0133] 在块 908,识别或产生在拍速提取窗期间出现的音符开始事件组。在某些实施例中,也识别或产生在拍速提取窗期间出现的休止符开始位置组。在块 910,提取音符开始间隔。音符开始间隔表示在每个音符或休止符的开始与随后的音符或休止符的开始之间逝去

的时间量。如上所述,音符开始间隔可与音符时值相同或不同。

[0134] 通过确定每个被提取的音符开始间隔相对于在块 904 确定的理想化音符值的误差值,方法 240 在块 902 继续。在一个实施例中,每个音符开始间隔在块 922 被每个参考音符时值除。结果可接着用于在块 924 确定与音符开始间隔最接近的参考音符时值(或参考音符时值的倍数)。

[0135] 例如,音符开始间隔可为 35,650 个样本。由不同的参考音符时值除音符开始间隔以及取差值的绝对值可产生不同的结果,每个结果表示误差值。例如,该音符开始间隔与 72bpm(36,750 个样本)的参考四分音符比较的误差值可为大约 0.03,而该音符开始间隔与 76bpm(17,408 个样本)的参考八分音符比较的误差值可为大约 1.05。最小误差值可接着用于确定最接近的参考音符时值(例如,在本示例性情况下,在 72bpm 的四分音符)。

[0136] 在一些实施例中,在多个音符开始事件的范围中产生一个或多个误差值。在一个实施例中,在确定最小复合误差值之前,在数学上合并拍速提取窗中的所有音符开始事件的误差值。例如,可合计、平均或以在数学上合并不同音符开始事件的误差值。

[0137] 一旦在块 920 确定了误差值,就在块 930 确定最小误差值。与最小误差值相关的参考拍速可接着用作所提取的拍速。在上面的例子中,最低误差值从 72bpm 的四分音符的参考音符时值产生。因此,72bpm 可被确定为在给定窗内的提取的拍速。

[0138] 一旦确定了拍速,就可能希望为音频输入信号中(或至少在信号的窗中)识别出的每个音符或休止符分配音符值。图 10 提供了根据本发明实施例的用于确定音符值的示例性方法的流程图。方法 245 通过确定在图 9 的块 930 提取的拍速的第二组参考音符时值而在块 1002 开始。在一些实施例中,第二组参考音符时值与第一组参考音符时值相同。在这些实施例中,应认识到,第二组可被简单地被提取为第一组参考音符时值的子集。在其他实施例中,第一组参考音符时值只包括可能的音符值的子集,而第二组参考音符时值包括提取的拍速的可能音符时值的更完整的集合。

[0139] 在块 1004,方法 245 可针对窗口中的从音频输入信号提取的音符开始事件产生或识别所接收的音符时值。所接收的音符时值可表示出现在窗口期间的音符和休止符的实际时值,与第二组参考音符时值所表示的理想化时值相对。在块 1006,所接收的音符时值与参考音符时值比较,以确定最接近的参考音符时值(或参考音符时值的倍数)。

[0140] 最接近的参考音符时值可接着被分配到音符或休止符作为其音符值。在一个例子中,所接收的音符时值被确定为大约 1.01 个参考四分音符,并可被分配一个四分音符的音符值。在另一例子中,所接收的音符时值被确定为大约 1.51 个参考八分音符,并可被分配一个附点八分音符(或用连接线连到十六分音符的八分音符)的音符值。

[0141] 图 12 提供了示出该示例性拍速检测方法的示例性数据的曲线。曲线 1200 绘制出复合误差值与以每分钟拍为单位的拍速的关系曲线。方块点 1202 表示由于使用参考四分音符的误差值,而菱形点 1204 表示由于使用参考八分音符的误差值。例如,曲线 1200 上的第一方块点 1202-1 示出对于与在 72bpm 的参考四分音符比较的一组音符开始间隔,产生大约 3.3 的误差值。

[0142] 曲线 1200 示出,四分音符参考时值 1210-1 的最小误差和八分音符参考时值 1210-2 的最小误差都在 84bpm 产生。这可指示在音频输入信号的窗内,提取的拍速是 84bpm。



[0143] 图 11 提供了示出图 12 所示的示例性拍速检测方法的额外的示例性数据。该组音符开始间隔 1102 的一部分被示出,以范围从 7,881 到 63,012 个样本的样本数量被测量。对照一组参考音符时值 1104 来评估音符开始间隔 1102。如所示,参考音符时值 1104 包括在 8 个参考拍速上 4 个音符值的以秒和样本为单位的时值(假定采样率为每秒 44,100 个样本)。如图 12 所示,确定提取的拍速为 84bpm。与 84bpm 的参考拍速有关的参考音符时值 1106 被提取,并与音符开始间隔比较。最接近的参考音符时值 1108 被识别出。这些时值可接着用于将音符值 1110 分配给每个音符开始间隔(或在每个音符开始间隔处开始的每个音符的时值)。

#### [0144] 音调检测

[0145] 确定音频输入信号的一部分的音调可能对产生有用的总谱输出很重要。例如,确定音调可为乐曲的部分提供调号,并可识别应使用标识音符的地方。然而,确定音调可能由于多种原因很难。

[0146] 一个原因是乐曲常常在音调之间移动(例如,通过变调)。例如,摇滚歌曲可能有在 G 大调的音调中的独唱部,对每次合唱变调到 C 大调的音调,并在过渡乐节进一步变调到 D 小调。另一原因是乐曲常常包含多个临时符号(不“在音调中”的音符)。例如,以 C 大调的歌曲(其不包含升号或降号)可使用升号或降号来给音符短句添加色彩或张力。又一原因是乐曲在音调之间常常有过渡时期,其中短句展示混合音调的性质。在这些混合状态中,确定音调何时变化或音乐的哪些部分属于哪个音调可能很难。例如,在从 C 大调到 F 大调的过渡乐节,歌曲可反复地使用 B- 降号。这将表现为在 C 大调的音调中而不是在 F 的音调中的临时符号。因此,可能希望确定音调变化在哪里出现,使得音乐总谱表示 170 不会不正确地反映临时符号或在音调之间反复地摇摆。确定音调可能很难的再一原因是,多个音调可能有相同的调号。例如,在 C 大调、A 小调或 D 多利安(dorian)的任何一个中没有升号或降号。

[0147] 图 13 提供了根据本发明实施例的用于检测音调的示例性方法的流程图。方法 250 通过在块 1302 确定一组音调代价函数而开始。代价函数可例如试图在特定的时间窗内最小化在一首音乐中的临时符号的数量。

[0148] 图 14A 和 14B 提供了根据本发明实施例的用在音调检测中的两个示例性音调代价函数的图示。在图 14A 中,音调代价函数 1400 基于在不同音调中的一系列自然音阶(diatonic scale)。对该音调的自然音阶中的所有音符给出“1”的值,而对不在该音调的自然音阶中的所有音符给出“0”的值。例如,C 大调的音调包含下列自然音阶:C-D-E-F-G-A-B。因此,代价函数 1400 的第一行 1402-1 只对那些音符显示“1”。

[0149] 在图 14B 中,音调代价函数 1450 也基于在不同音调中的一系列自然音阶。与图 14A 中的代价函数 1400 不同,图 14B 中的代价函数 1450 为给定音调中的所有第一、第三和第五音阶音分配“2”的值。尽管如此,对该音调的自然音阶中的所有其他音符给出“1”的值,而对不在该音调的自然音阶中的所有音符给出“0”的值。例如,C 大调的音调包含自然音阶 C-D-E-F-G-A-B,其中第一音阶音是 C,第三音阶音是 E,且第五音阶音是 G。因此,代价函数 1450 的第一行 1452-1 显示 2-0-1-0-2-1-0-2-0-1-0-1。

[0150] 该代价函数 1450 可能由于很多原因是有用的。一个原因是,在很多音乐类型(例如,民谣、摇滚、古典等)中,第一、第三和第五音阶音往往在听者中产生某个音调的感觉时

有心理声学意义。因此,对这些音符加大代价函数的权重,在某些情况下可提高音调确定的准确性。使用该代价函数 1450 的另一原因可以是区分具有相似的调号的音调。例如,C 大调、D 多利安、G 混合里第亚 (mixolydian)、A 小调和其他音调都不包含升号或降号。然而,这些音调中的每个都有彼此不同的第一、第三和 / 或第五音阶音。因此,在音阶中的所有音符的相等权重可展现在这些音调的存在之间的小差异 (即使可能有明显的心理声学差异),但经调节的权重可改进音调确定。

[0151] 应认识到,可为了不同的原因对代价函数进行其他调节。在一个实施例中,代价函数可被不同地加以权重,以反映音频输入信号 (例如,从用户接收,来自音频文件中的头部信息,等等) 的种类。例如,蓝调代价函数可根据五音阶而不是自然音阶、音调音阶对音符更大地加以权重。

[0152] 返回到图 13,在块 1304 可确定音调提取窗。音调提取窗可为跨越音频输入信号的某个连续部分的预先确定的或自适应的时间窗。优选地,音调提取窗宽到足以覆盖大量音符开始事件。因此,块 1304 的某些实施例调节拍速提取窗的宽度,以覆盖预定数量的音符开始事件。

[0153] 在块 1306,识别或产生在音调提取窗期间出现的一组音符开始事件。接着,在块 1308 确定每个音符开始事件的音符音高。可在块 1308 以任何有效的方式确定音符音高,包括通过上面描述的音高确定方法。应认识到,因为音符开始事件表示时间位置,在技术上在该时间位置不能有音高 (音高确定需要某个持续时间)。因此,在音符开始时的音高通常指与跟随音符开始事件的音符时值相关的音高。

[0154] 在块 1310,可对照每个代价函数评估每个音符音高,以产生一组误差值。例如,假定对于音频输入信号的窗,音符音高的序列如下 :C-C-G-G-A-A-G-F-F-E-E-D-D-C。对照图 14A 中的代价函数 1400 的第一行 1402-1 评估该序列,可产生  $1+1+1+1+1+1+1+1+1+1+1+1+1+1 = 14$  的误差值。对照图 14A 中的代价函数 1400 的第三行 1402-2 评估该序列可产生  $0+0+1+1+1+1+1+0+0+1+1+1+1+0 = 9$  的误差值。重要地,对照图 14A 中的代价函数 1400 的第四行 1402-3 评估该序列可产生与当使用第一行 1402-1 时相同的 14 的误差值。使用该数据,音高序列在 D 大调的音调中看起来相对不可能,但不可能确定是 C 大调还是 A 小调 (其共享相同的调号) 是更可能的候选调号。

[0155] 使用图 14B 中的代价函数 1450 产生不同的结果。对照第一行 1452-1 评估该序列可产生  $2+2+2+2+1+1+2+1+1+2+2+1+1+2 = 22$  的误差值。对照第三行 1452-2 评估该序列可产生  $0+0+1+1+2+2+1+0+0+2+2+1+1+0 = 13$  的误差值。重要地,对照第四行 1452-3 评估该序列可产生  $2+2+1+1+2+2+1+1+1+2+2+1+1+2 = 21$  的误差值,其比当使用第一行 1452-1 时获得的误差值 22 少一。使用该数据,音高序列在 D 大调的音调中看起来 相对不可能,但该序列在 C 大调中而不是在 A 小调中现在看起来稍微更可能。

[0156] 应认识到,由于非零值被分配给音调内的音符的事实,当所接收的音符更可能在给定的音调中时,上面讨论的代价函数 (例如 1400 和 1450) 产生较高的结果。然而其他实施例根据代价函数的标准可将“0”分配给“大部分在音调中”的音高。使用代价函数的这些其他实施例可为较不匹配的音调产生较大的数字,从而产生可能更直观的误差值 (即,较大的误差值表示较差的匹配)。

[0157] 在块 1312,不同音调代价函数的不同误差值被比较以产生与音符音高序列最佳匹

配的音调。如上所述,在一些实施例中,这可包括找到最高的结果(即,最佳匹配),而在其他实施例中,这可包括找到最低的结果(即,最不匹配的误差),这取决于代价函数的构成。

[0158] 值得注意的是,根据本发明音调确定的其他方法是可能的。在一些实施例中,人工神经网络可用于进行或改进复杂的音调确定。在其他实施例中,可对照代价函数评估音调变化的序列,以改进音调确定。例如,方法 250 可检测模式 C 大调 -F 大调 -G 大调 -C 大调的音频输入信号中的一系列音调。然而,由于检测到很多 B- 本位音(F 的升号 -4- 在大多数音乐种类中不可能的音符),可限制检测到 F 大调的置信度。考虑到被识别为 F 大调的音调先于以 C 大调开始和结束的歌曲的 G 大调中的部分,即使偶尔的 B- 本位音的存在也可指示音调确定应被修改成更相称的选择(例如, D 多利安或甚至 D 小调)。

[0159] 一旦确定了音调,就可能希望使音调音高标志符合在每个音符开始事件(至少对于出现在音调提取窗内的那些开始事件)的音符。图 15 提供了根据本发明实施例的用于确定音调音高标志的示例性方法的流程图。方法 255 通过在块 1502 产生提取的音调的一组参考音高而开始。

[0160] 值得注意的是,可能的音高对所有音调(例如,特别考虑现代调谐标准)可能是相同的。例如,钢琴的每个八度中的所有 12 个半音音符可在任何音调中表演。差异可以是那些音高如何在总谱上表示(例如,不同的音调可给相同的音符音高分配不同的临时符号)。例如,以 C 大调的钢琴上的“白色键”的音调音高可被指定为 C、D、E、F、G、A 和 B。以 D 大调的相同组的音调音高可被指定为 C- 本位音、D、E、F- 本位音、G、A 和 B。

[0161] 在块 1504,与每个提取的音符音高最接近的参考音高被确定并用于产生该音符的音调音高确定。音调音高确定可接着在块 1506 被指定给音符(或音符开始事件)。

#### [0162] 示例性硬件系统

[0163] 上面描述的系统和方法可用很多方法实现。一个这样的实现包括各种电子部件。例如,使用适合于在硬件中执行一些或全部适用功能的一个或多个专用集成电路(ASIC)可单独地或共同地实现图 1B 中的系统的单元。可选地,功能可由一个或多个其他处理单元(或核心)在一个或多个集成电路上执行。在其他实施例中,可使用其他类型的集成电路(例如,结构/平台 ASIC、现场可编程门阵列(FPGA)和其他半定制 IC),其可用本领域已知的任何方式编程。每个单元的功能也可使用包含在存储器中的指令的整体或部分来实现,这些指令被设计成由一个或多个通用或专用处理器执行。

[0164] 图 16 提供了用于实现本发明的某些实施例的计算系统 1600 的结构图。在一个实施例中,计算系统 1600 可起图 1A 所示的系统 100 的作用。应注意,图 16 只表示提供各种部件的一般化图示,在适合时可利用这些部件中的任何一个或全部。因此,图 16 广泛示出各个系统元件可如何以相对分开的或相对更综合的方式实现。

[0165] 示出包括硬件元件的计算系统 1600,这些硬件元件可通过总线 1626 电耦合(或视情况以其他方式进行通信)。硬件元件可包括:一个或多个处理器 1602,其包括但不限于一个或多个通用处理器和/或一个或多个专用处理器(例如,数字信号处理芯片、图形加速芯片和/或类似物);一个或多个输入设备 1604,其可包括但不限于鼠标、键盘和/或类似物;以及一个或多个输出设备 1606,其可包括但不限于显示设备、打印机和/或类似物。

[0166] 计算系统 1600 可还包括(和/或连接到)一个或多个存储设备 1608,其可包括但不限于本地和/或网络可访问的存储器,和/或可包括但不限于磁盘驱动器、驱动器阵列、

光学存储设备、固态存储设备例如随机存取存储器 (RAM) 和 / 或只读存储器 (ROM), 这些存储器可为可编程的、可闪速更新的, 等等。计算系统 1600 也可包括通信子系统 1614, 其可包括但不限于调制解调器、网卡 (无线或有线)、红外通信设备、无线通信设备和 / 或芯片组 (例如蓝牙设备、802.11 设备、WiFi 设备、WiMax 设备、蜂窝通信设施等) 和 / 或类似物。通信子系统 1614 可允许使用网络 (例如, 下面描述的网络, 列举一个例子) 和 / 或这里描述的任何其他设备交换数据。在很多实施例中, 计算系统 1600 将还包括工作存储器 1618, 其可包括 RAM 或 ROM 设备, 如上所述。

[0167] 计算系统 1600 也可包括被显示为当前位于工作存储器 1618 内的软件元件, 工作存储器 1618 包括操作系统 1624 和 / 或其他代码, 例如一个或多个应用程序 1622, 应用程序 1622 可包括本发明的计算机程序, 和 / 或可设计成实现本发明的方法和 / 或配置本发明的系统, 如这里所述的。仅仅作为例子, 关于上面讨论的方法描述的一个或多个程序可被实现为计算机 (和 / 或计算机内的处理器) 可执行的代码和 / 或指令。这些指令和 / 或代码的集合可存储在计算机可读存储介质 1610b 上。在一些实施例中, 计算机可读存储介质 1610b 是上面描述的存储设备 1608。在其他实施例中, 计算机可读存储介质 1610b 可包括在计算机系统内。在又一些其他实施例中, 计算机可读存储介质 1610b 可与计算机系统分离 (即, 可移除的介质, 例如光盘等) 和 / 或设置在安装包中, 使得存储介质可用于使用存储在其上的指令 / 代码给通用计算机编程。这些指令可采取可由计算系统 1600 执行的可执行代码的形式, 和 / 或可采取源代码和 / 或可安装的代码的形式, 源代码和 / 或可安装的代码当在计算系统 1600 上编译和 / 或安装 (例如, 通过使用各种通常可得到的编译器、安装程序、压缩 / 解压缩应用程序等中的任何一种) 时接着呈现可执行代码的形式。在这些实施例中, 计算机可读存储介质 1610b 可由计算机可读存储介质阅读器 1610a 读取。

[0168] 对本领域技术人员来说很明显, 可根据特定的要求进行相当多的变化。例如, 也可使用定制的硬件, 和 / 或特定的元件可在硬件、软件 (包括可移植软件, 例如 Java 程序等) 或两者中实现。进一步地, 可使用到其他计算设备例如网络输入 / 输出设备的连接。

[0169] 在一些实施例中, 一个或多个输入设备 1604 可与音频接口 1630 耦合。音频接口 1630 可配置成与麦克风、乐器、数字音频设备或其他音频信号或文件源例如物理地、光学地、电磁地等连接。进一步地, 在一些实施例中, 一个或多个输出设备 1606 可与源转录接口 1632 耦合。源转录接口 1632 可配置成将由本发明的实施例产生的音乐总谱表示数据输出到能够处理该数据的一个或多个系统。例如, 源转录接口可配置成与总谱转录软件、总谱公布系统、扬声器等连接。

[0170] 在一个实施例中, 本发明使用计算机系统 (例如计算系统 1600) 来执行本发明的方法。根据一组实施例, 这样的方法的一些或全部程序由计算系统 1600 响应于处理器 1602 来执行, 处理器 1602 执行包含在工作存储器 1618 中的一个或多个指令的一个或多个序列 (所述指令可合并到操作系统 1624 和 / 或其他代码中, 例如应用程序 1622)。这样的指令可从另一机器可读介质被读取到工作存储器 1618 中, 所述另一机器可读介质例如是一个或多个存储设备 1608 (或 1610)。仅作为例子, 包含在工作存储器 1618 中的指令的序列的执行可使处理器 1602 执行这里描述的方法的一个或多个程序。

[0171] 如这里使用的术语“机器可读介质”和“计算机可读介质”指参与提供数据的任何介质, 该数据使机器以特定的方式操作。在使用计算系统 1600 实现的实施例中, 各种机器

可读介质可包括向处理器 1602 提供用于执行的指令 / 代码, 和 / 或可用于储存和 / 或承载这样的指令 / 代码 (例如, 作为信号)。在很多实现中, 计算机可读介质是物理和 / 或有形的存储介质。这样的介质可采取很多形式, 包括但不限于非易失性介质、易失性介质和传输介质。非易失性介质包括例如光盘或磁盘, 例如存储设备 (1608 或 1610)。易失性介质包括但不限于动态存储器, 例如工作存储器 1618。传输介质包括同轴电缆、铜线和光纤, 包括构成总线 1626 的导线以及通信子系统 1614 的不同部件 (和 / 或通信子系统 1614 提供与其他设备的通信所用的介质)。因此, 传输介质也可采取波的形式 (包括但不限于, 无线电波、声波和 / 或光波, 例如在无线电波和红外数据传输期间产生的波)。

[0172] 物理和 / 或有形的计算机可读介质的一般形式包括例如软盘、软碟、硬盘、磁带或任何其他磁性介质、CD-ROM、任何其他光学介质、穿孔卡片、纸带、具有孔的图案的任何其他物理介质、RAM、PROM、EPROM、FLASH-EPROM、任何其他存储器芯片或盒式磁带、诸如以下描述的载波、或计算机可从其读取指令和 / 或代码的任何其他介质。

[0173] 机器可读介质的各种形式可涉及将一个或多个指令的一个或多个序列传送到处理器 1602 用于执行。仅作为例子, 指令可最初被承载在远程计算机的磁盘和 / 或光盘上。远程计算机可将指令装入其动态存储器中, 并在传输介质上作为信号发送指令, 以被计算机系统 1600 接收和 / 或执行。可能以电磁信号、声信号、光信号和 / 或类似信号的形式存在的这些信号根据本发明的不同实施例都是载波的例子, 指令可编码在载波上。

[0174] 通信子系统 1614 (和 / 或其部件) 通常接收信号, 且总线 1626 接着可能将该信号 (和 / 或信号所承载的数据、指令等) 传送到工作存储器 1618, 处理器 1602 从工作存储器 1618 取回并执行指令。工作存储器 1618 所接收的指令可在被处理器 1602 执行之前或之后可选地存储在存储设备 1608 上。

#### [0175] 其他功能

[0176] 应认识到, 除了上面描述的功能以外, 很多其他处理功能也是可能的。一组额外的处理功能包括增加提供给用户的可定制性的量。例如, 实施例可允许本发明的各种部件和方法的增强的可定制性。

[0177] 在一些实施例中, 由于各种原因, 部件和方法的不同阈值、窗和其他输入中的每个都是可调节的。例如, 如果对音调的确定看起来进行得太频繁 (例如, 用户可能不希望将从音调短暂的偏离作为音调变化展现在总谱上), 用户能够调节音调提取窗。再例如, 录制品可包括在录制时在表演期间使用的 60Hz 功率出现的背景噪声。用户可能希望调节各种滤波算法, 以忽略此 60Hz 音高, 以便不将它表示为总谱上的下部音符。在又一例子中, 用户可调节音高被量化的音乐盒的分辨率, 以调节音符音高分辨率。

[0178] 在其他实施例中, 可给用户提供更小的可定制性。在一个实施例中, 用户能够调节表示准确性水平。根据一个或多个参数, 包括选择对单独的总谱表示元素如拍速和音高之类的准确性, 用户可输入 (例如, 通过物理或虚拟滑块、旋钮、开关等) 系统是应产生更准确的还是较不准确的总谱表示。

[0179] 例如, 很多内部设置可一起起作用, 使得最小音符值是十六分音符。通过调节表示的准确性, 较长或较短的时值可被检测到并表示为最小值。这可能在表演者不严格按恒定拍子表演 (例如, 没有打击乐器组, 没有节拍器) 的情况下是有用的, 且太灵敏的系统可产生不合需要的表示 (例如, 三拍 附点音符)。作为另一例子, 很多内部设置可一起起作用,

以便最小音高变化是半音级（即，在半音音阶上的音符）。

[0180] 在又一些其他实施例中，可给用户甚至更少的可定制性。在一个实施例中，用户可输入他或她是新用户还是高级用户。在另一实施例中，用户可输入系统是应具有高灵敏性还是低灵敏性。在任一实施例中，在很多部件或方法中的很多不同的参数可在一起调节，以符合期望水平。例如，在一种情况下，歌手可能希望准确地转录音高和时值中的每个波动（例如，作为对找到错误或以所有其美学细微点忠实地再现特定的表演的实际帮助）；而在另一情况下，歌手可能希望通过使系统忽略小偏差来产生容易阅读的总谱用于发表。

[0181] 另一组额外的处理功能包括使用不同类型的输入来改进或以其他方式影响输入音频信号的处理。一个实施例使用一个或多个被训练的人工神经网络（ANN）来改进某些确定。例如，心理声学确定（例如，节拍、音调、乐器等）可完全适合于使用经训练的 ANN。

[0182] 另一实施例给用户使多个声道分层的功能（例如，一人乐队）。用户可通过表演鼓声道来开始，鼓声道使用本发明的系统被实时地处理。用户可接着相继地表演吉他声道、键盘声道和歌唱声道，其中每个都被处理。在一些情况下，用户可选择多个声道来一起处理，而在其他情况下，用户可选择使每个声道被单独地处理。来自一些声道的信息可接着用于改进或引导其他声道的处理。例如，鼓声道可被独立地处理，以产生高置信度拍速和节拍信息。该拍速和节拍信息可接着与其他声道一起使用来更准确地确定音符时值和音符值。又例如，吉他声道可在小时间窗内提供很多音高，这可使确定音调更容易。该音调确定可接着用于将音调音高确定分配给键盘声道中的音符。对于又一例子，多个声道可在一维或多维中被排列、量化或标准化（例如，声道可被标准化为具有相同的拍速、平均音量、音高音域、音高分辨率、最小音符时值等）。进一步地，在“一人乐队”的一些实施例中，用户可使用一个乐器来产生音频信号，接着使用本系统或方法来转换到不同的一个或多个乐器（例如，使用键盘表演四重奏的所有四个声道，并使用该系统来将键盘输入转换成弦乐四重奏）。在一些情况下，这可包括调节音色、使音乐谱线变调和其他处理。

[0183] 又一实施例使用音频输入信号非固有的输入来改进或引导处理。在一个实施例中，从用户、从另一系统（例如，计算机系统或互联网）或从数字音频文件中的头部信息来接收类别信息，以改进各种代价函数。例如，音调代价函数可对蓝调、印度古典、民谣等不同；或不同的乐器可能在不同的类别中更可能（例如，“风琴类”声音可能在赞美诗音乐中更可能是管风琴，而在波尔卡音乐中更可能是手风琴）。

[0184] 第三组额外的处理功能包括使用在多个部件或方法中的信息来改进复杂的确定。在一个实施例中，乐器识别方法的输出用于根据所识别的乐器的已知功能或限制来改进确定。例如，比如乐器识别方法确定音乐谱线可能正由钢琴表演。然而，音高识别方法确定音乐谱线包含快速微弱的颤音（例如，只在检测到的音调音高标志的一个或两个半音程内的音高的颤音）。因为这一般不可能是在钢琴上产生的效果，系统可确定谱线正由另一乐器（例如，电子键盘或风琴）表演。

[0185] 应认识到，根据本发明很多这样的额外处理功能是可能的。进一步地，应注意，上面讨论的方法、系统和设备只被确定为例子。必须强调，不同实施例可在适合时省略、替换或添加各种程序或部件。例如，应认识到，在可选实施例中，方法可按不同于所描述的顺序执行，且可添加、省略或合并不同的步骤。此外，关于某些实施例描述的特征可合并在各种其他实施例中。实施例的不同方面和元件可用类似的方式合并。此外，应强调技术是发展

的,因此这些元件中的很多仅是示例性的,不应被解释为限制本发明的范围。

[0186] 在说明书中给出了具体的细节,以提供对实施例的彻底理解。然而,本领域普通技术人员应理解,在没有这些具体细节的情况下可实现该实施例。例如,在没有不必要的细节的情况下示出了公知的电路、过程、算法、结构和技术,以便避免使实施例含糊。进一步地,这里提供的标题只用来有助于使不同实施例的描述的清楚,且不应被解释为限制本发明的范围或本发明的任何部分的功能。例如,某些方法或部件可被实现为其他方法或部件的部分,即使它们在不同的标题下被描述也是如此。

[0187] 此外,注意,实施例可被描述为一过程,其被描绘为流程图或结构图。虽然每个实施例可将操作描述为顺序的过程,但可并行或同时执行很多操作。此外,可重新排列操作的顺序。过程可以有不包括在附图中的额外步骤。

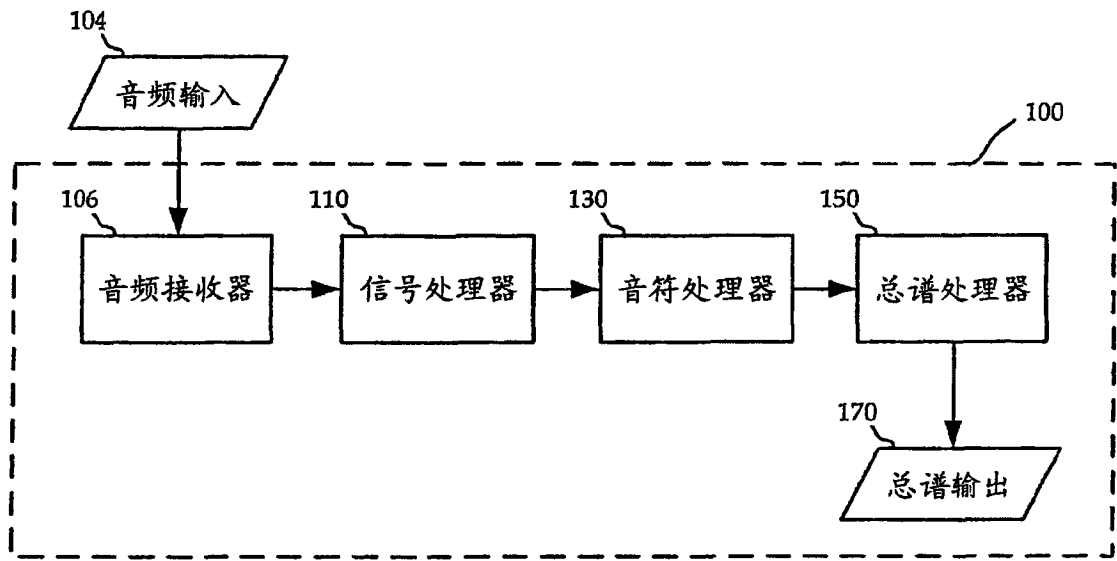


图 1A



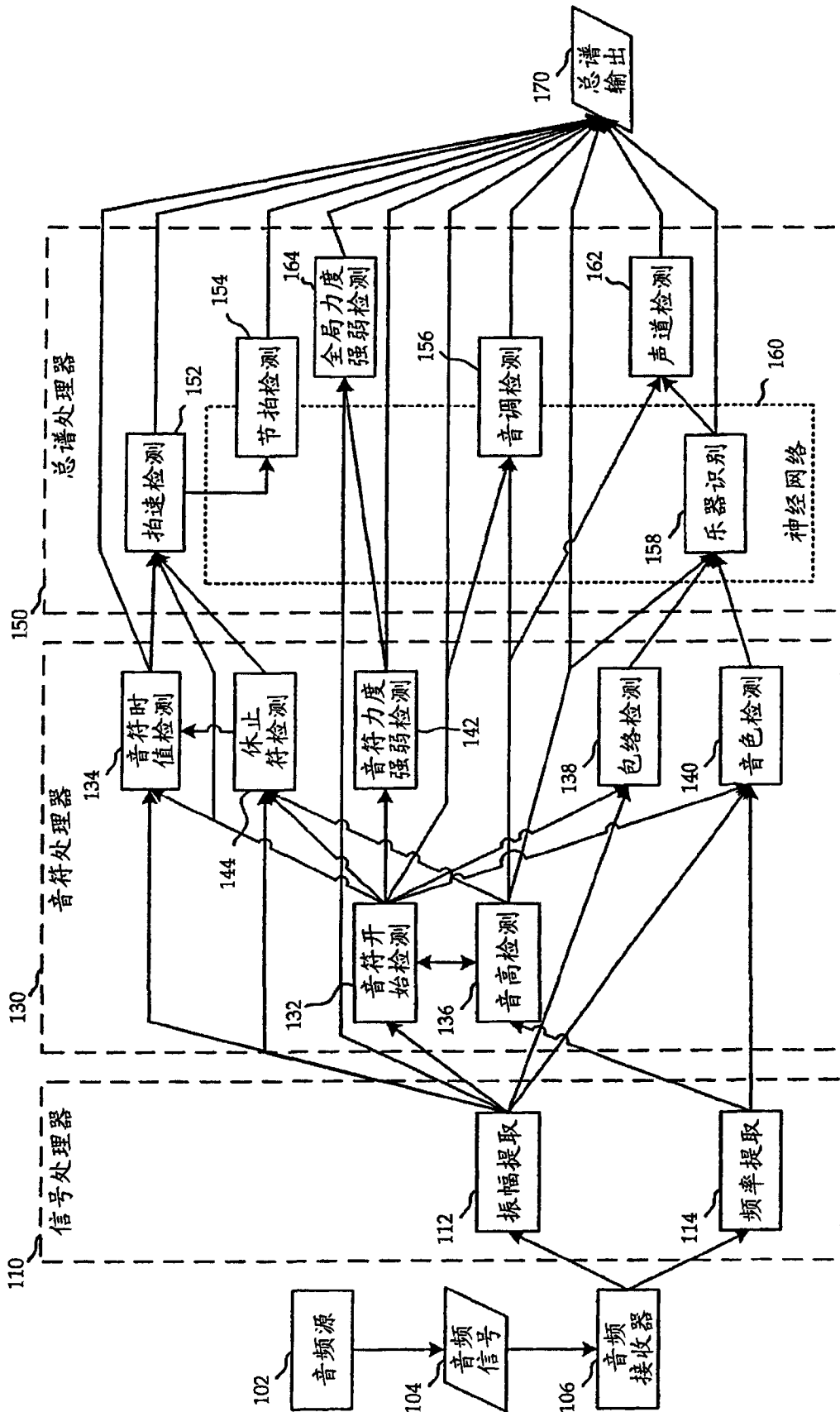


图 1B

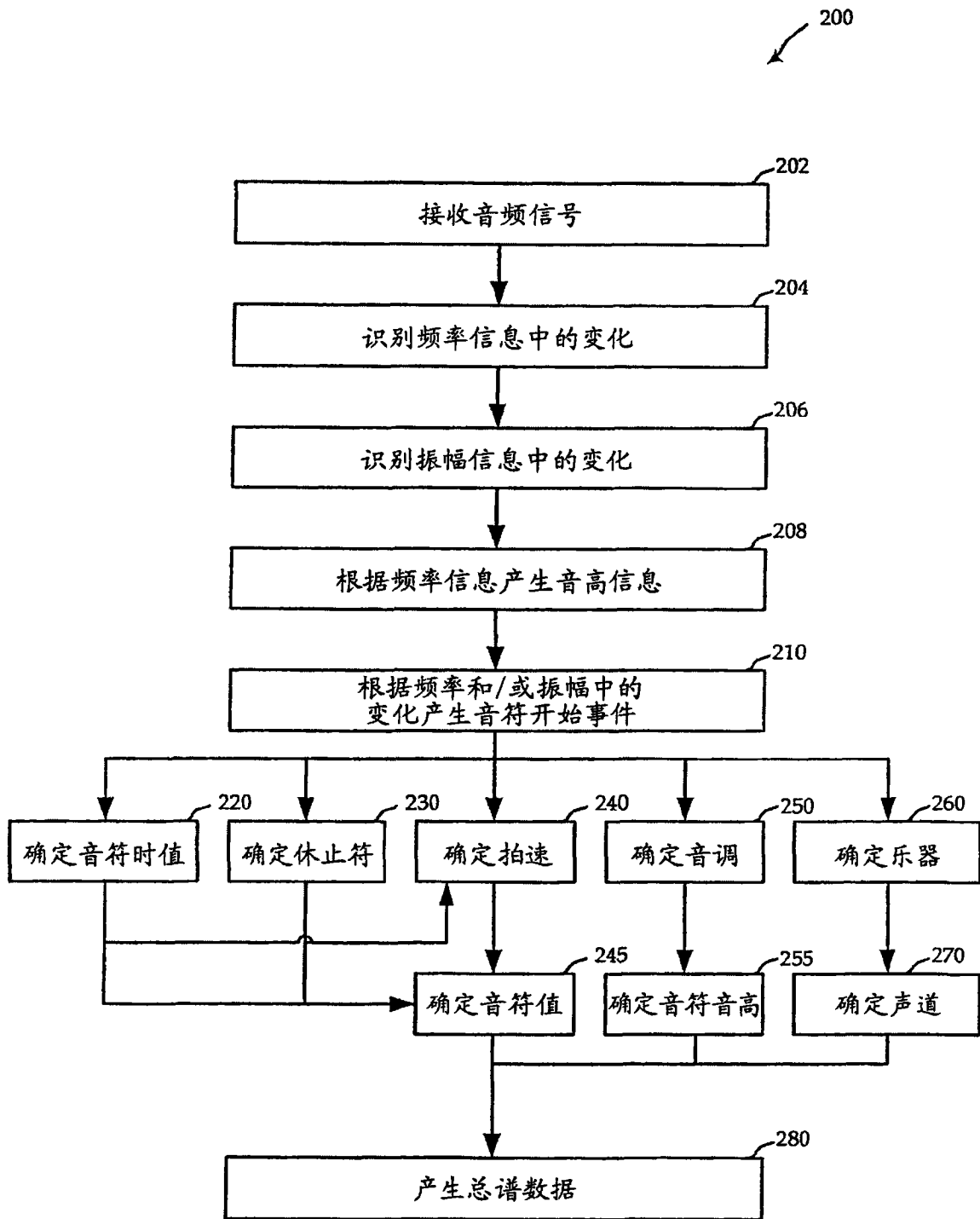


图 2

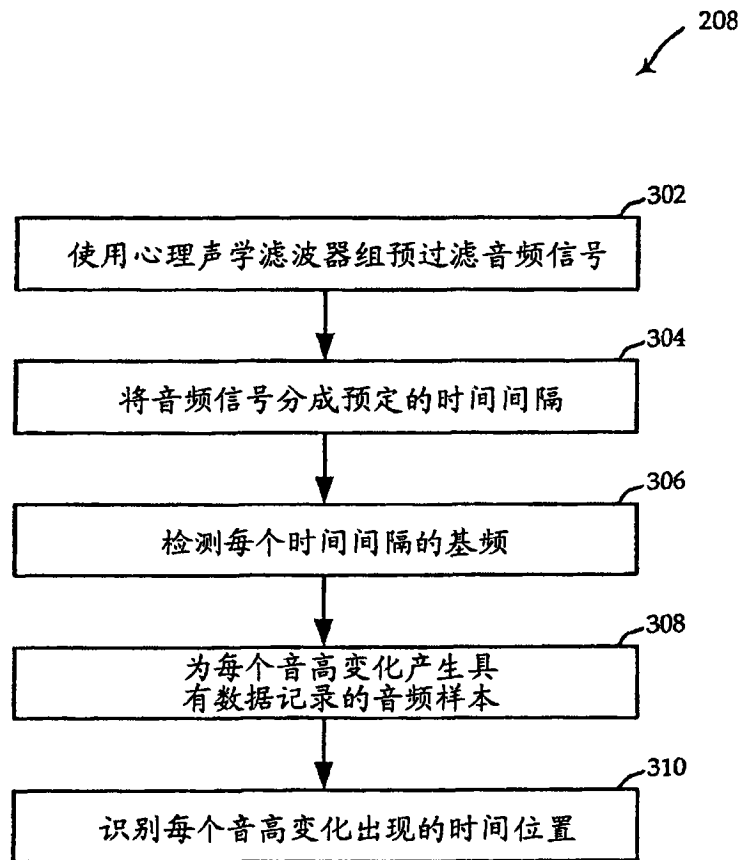


图 3

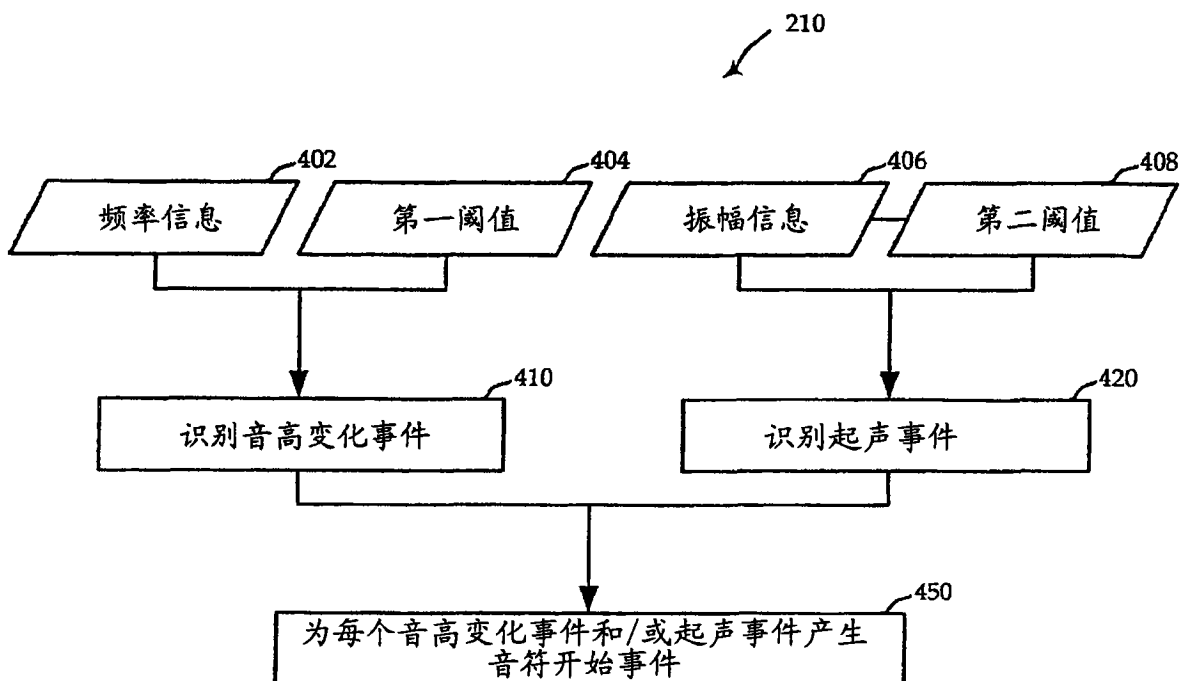


图 4A

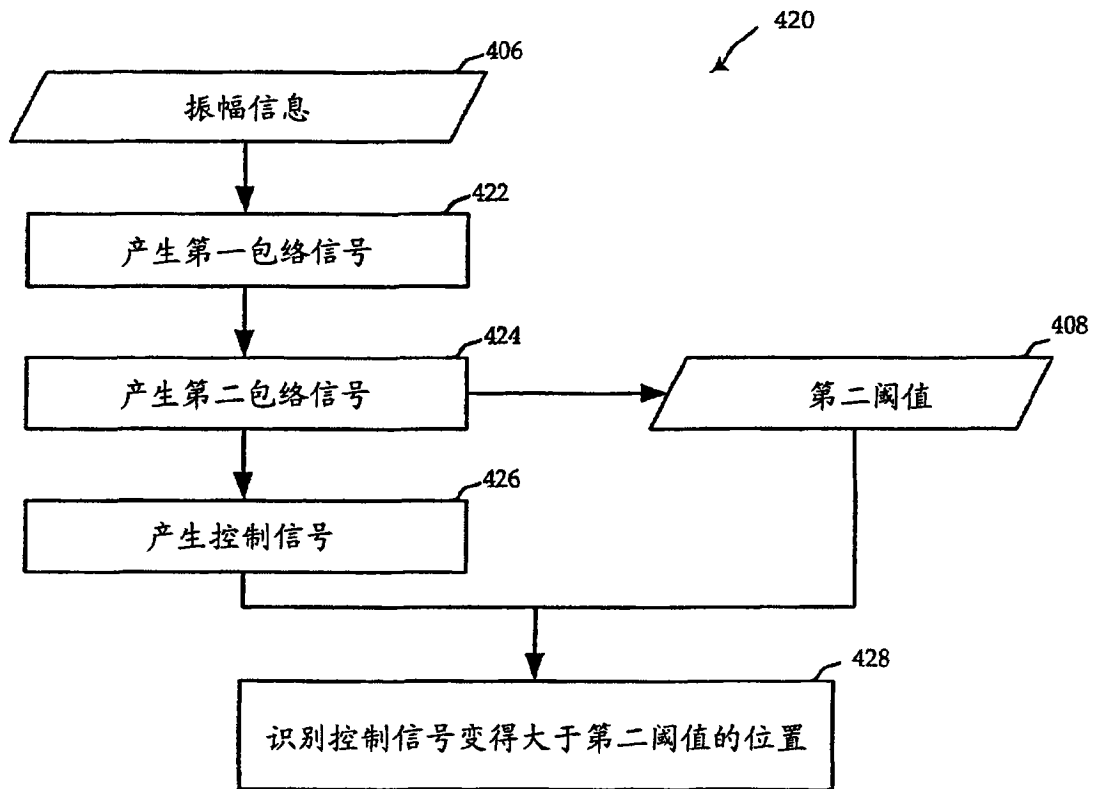


图 4B

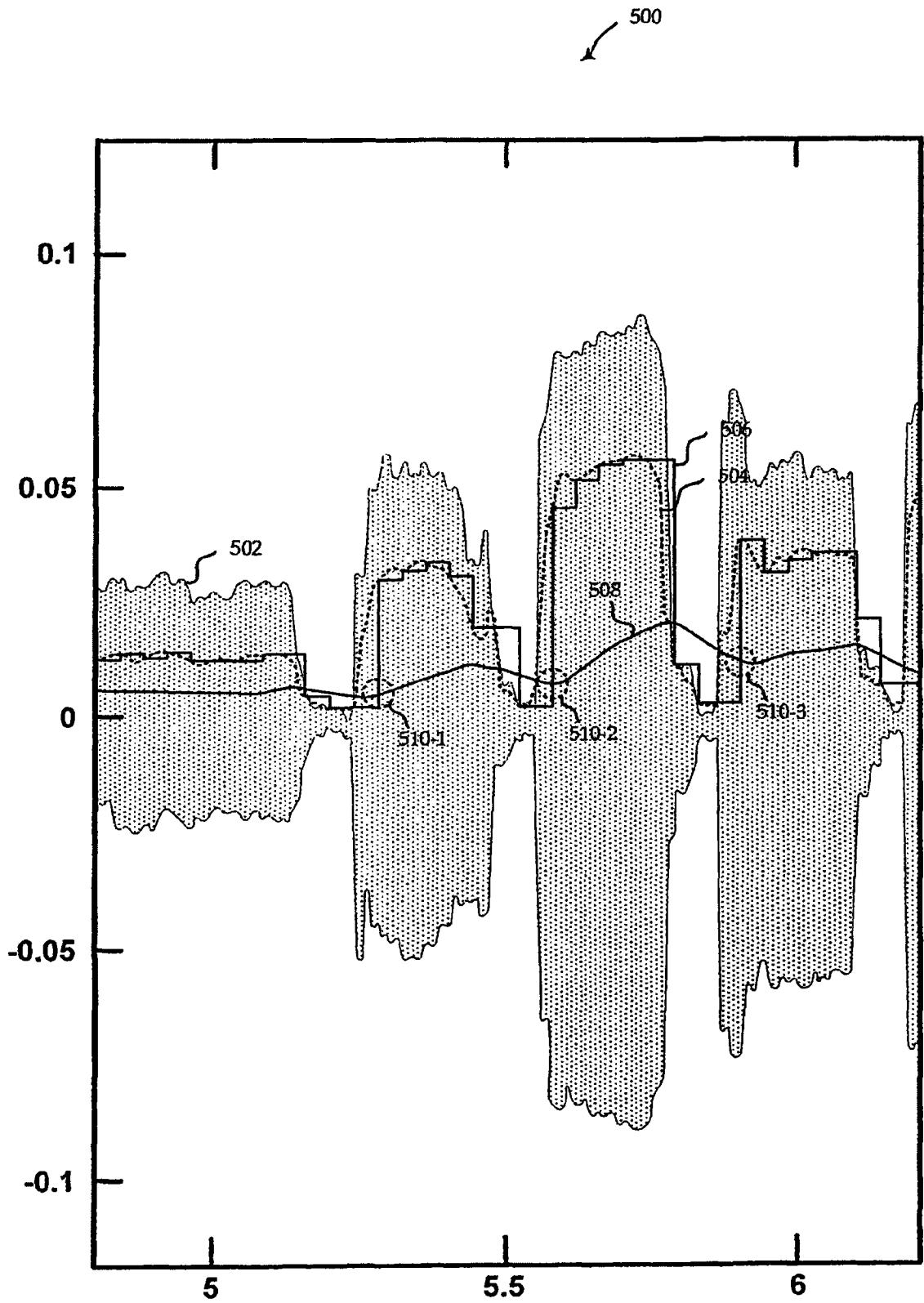


图 5

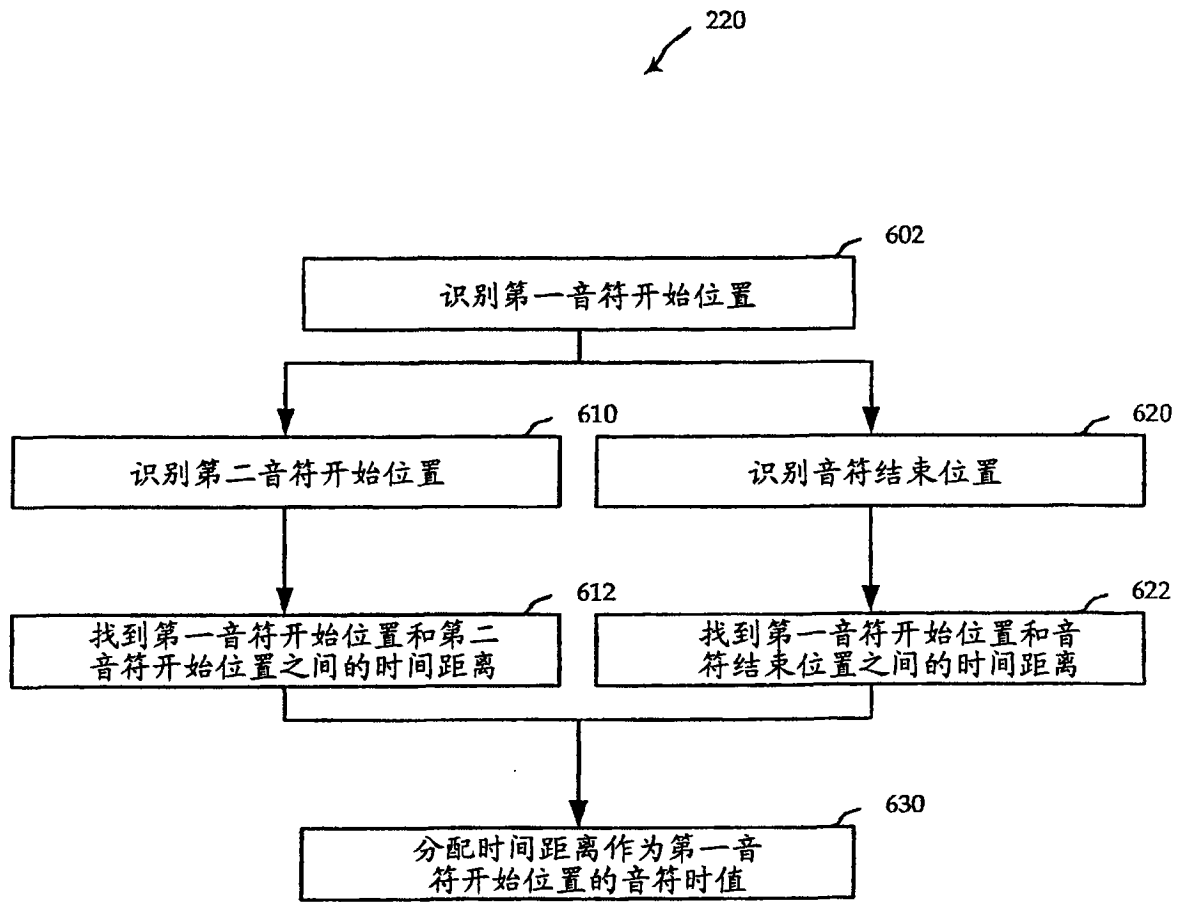


图 6

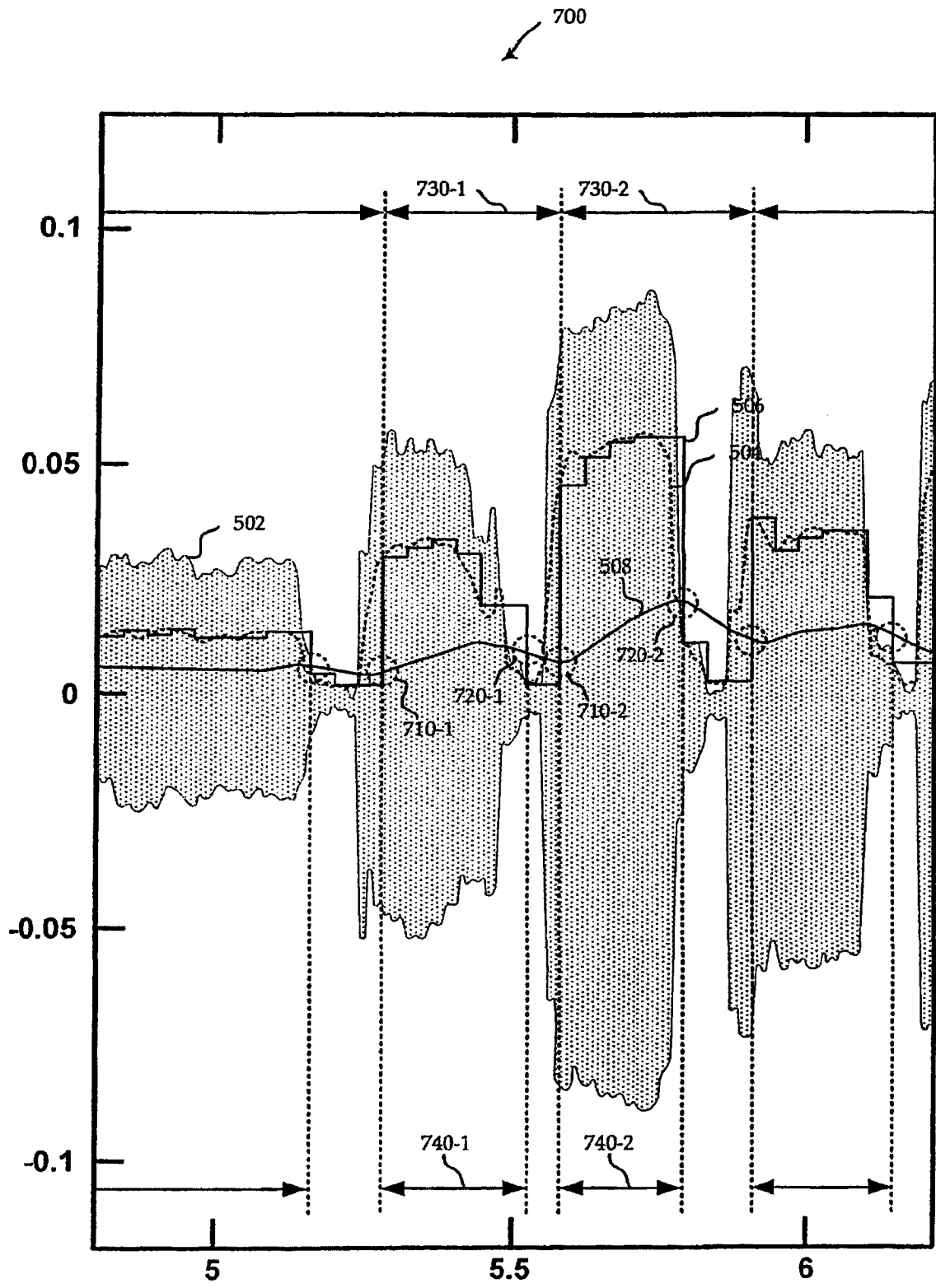


图 7

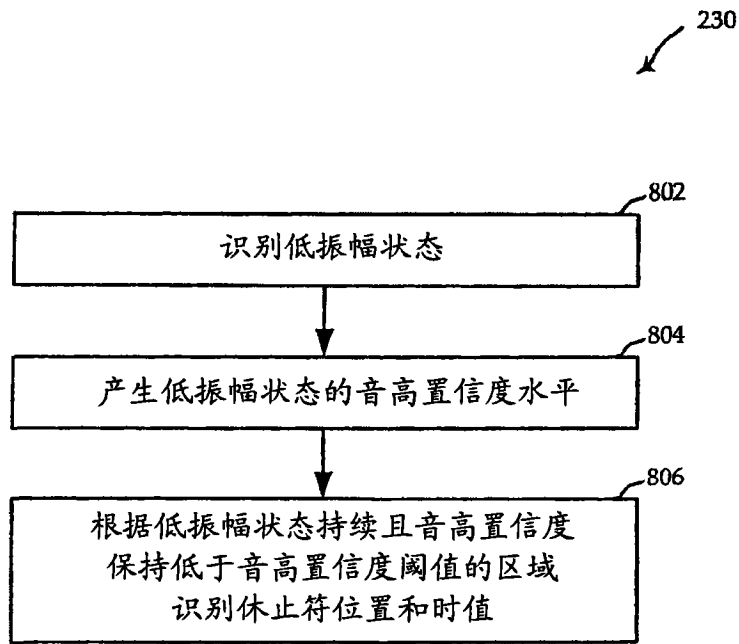


图 8



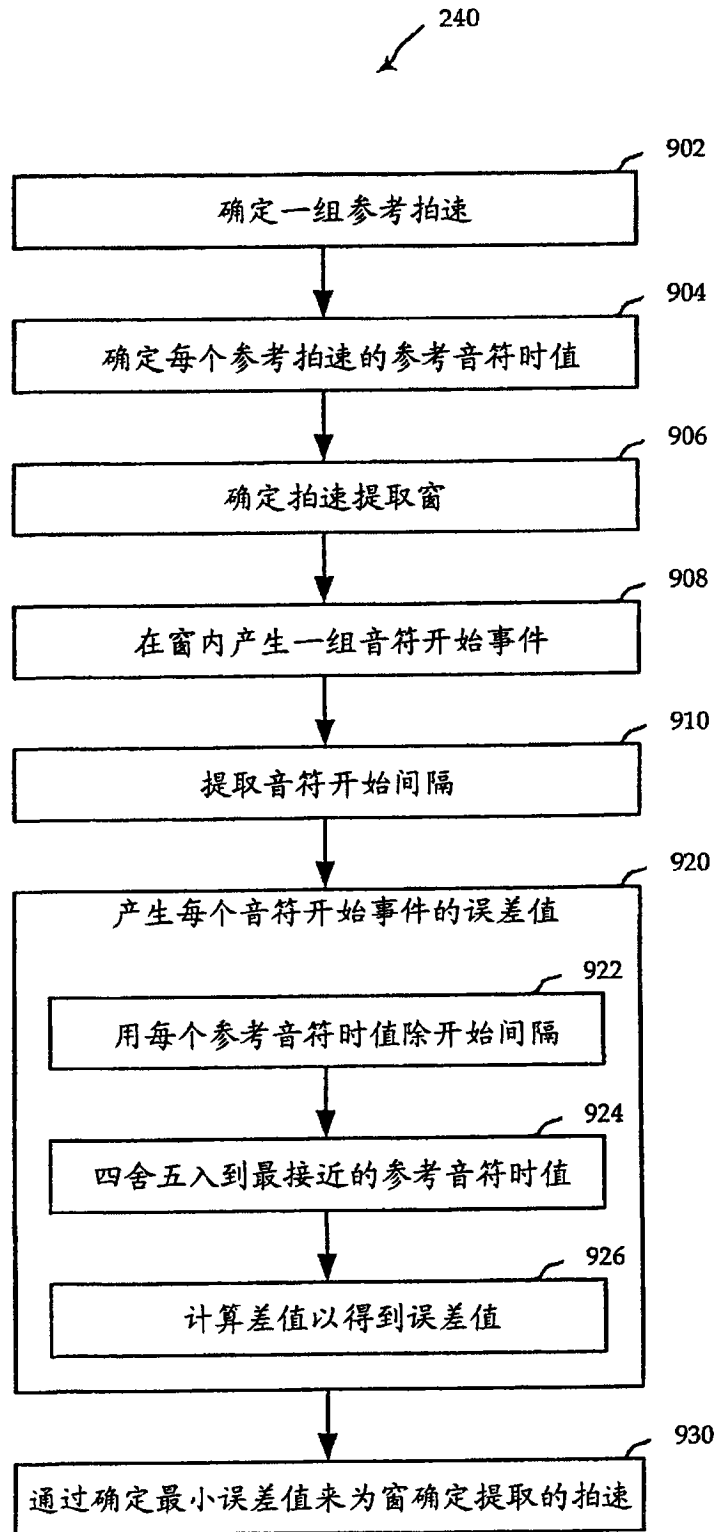


图 9

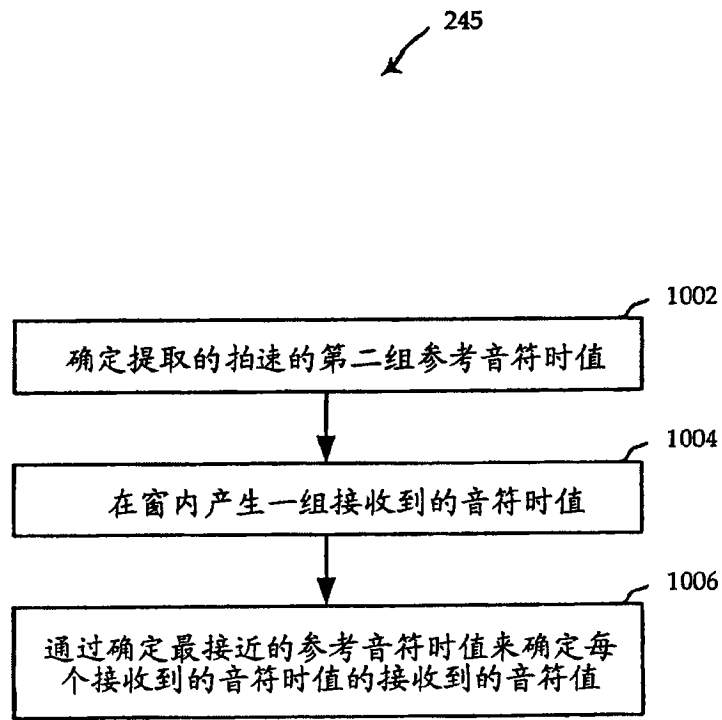


图 10

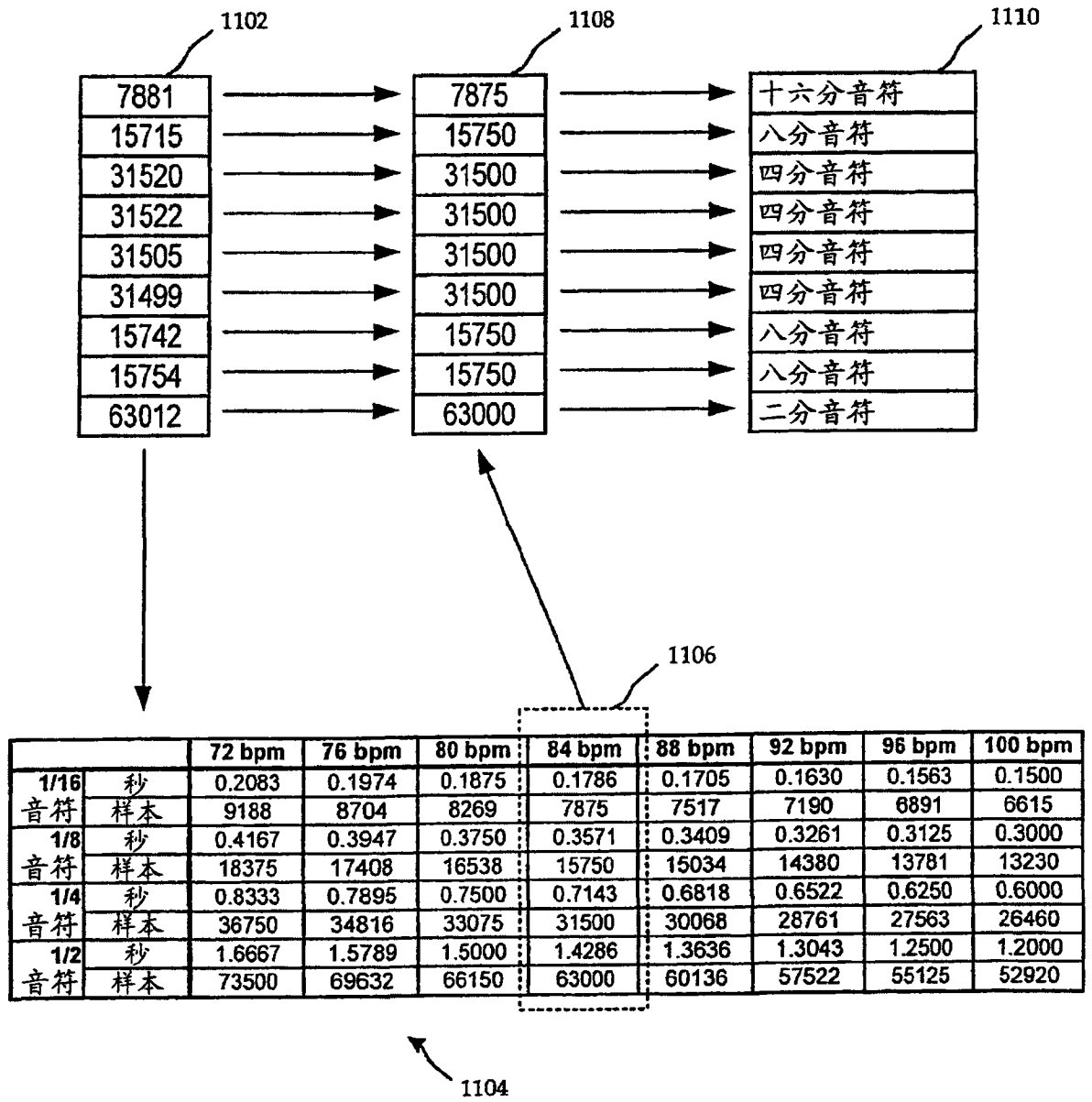


图 11

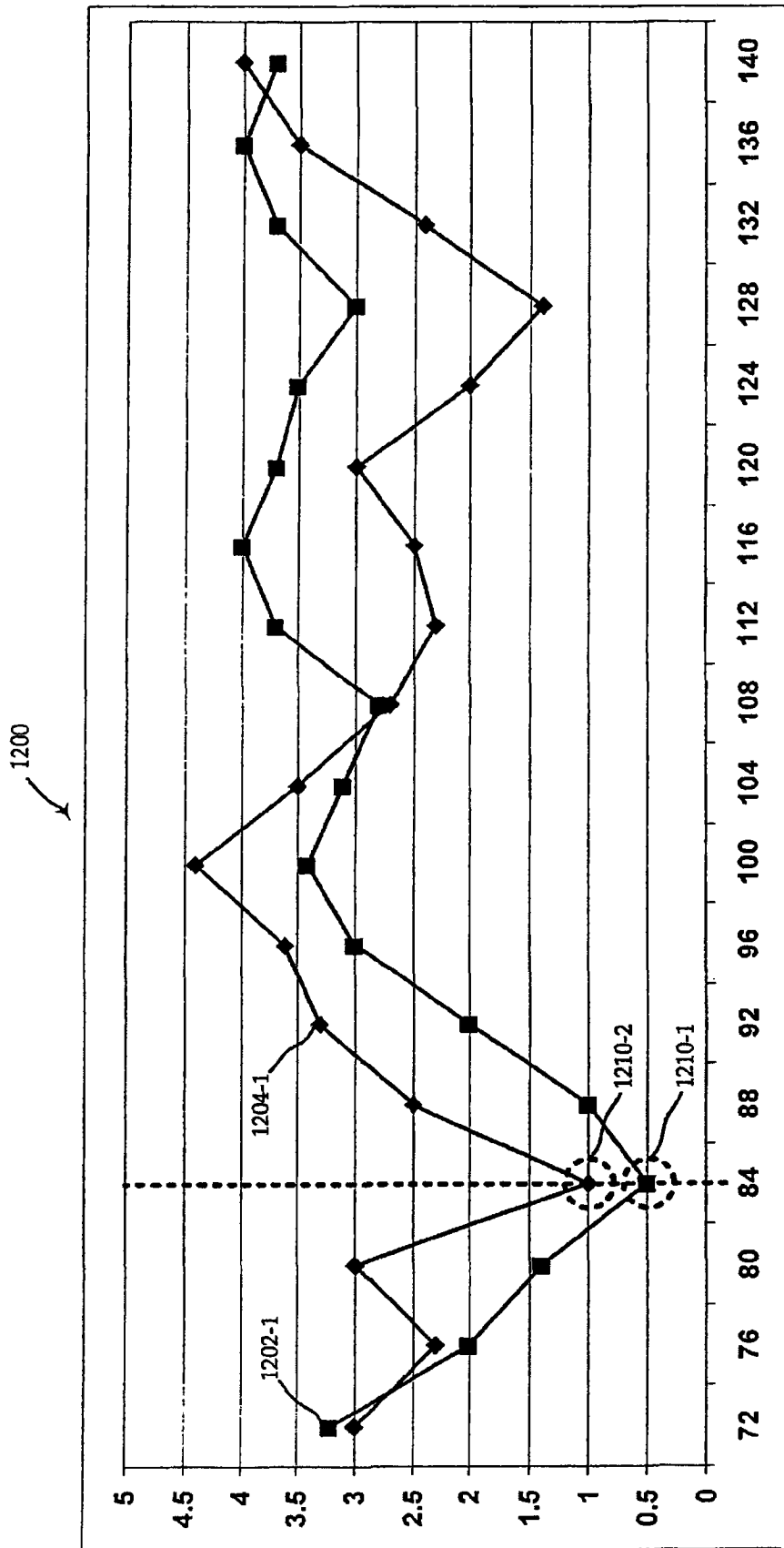


图 12

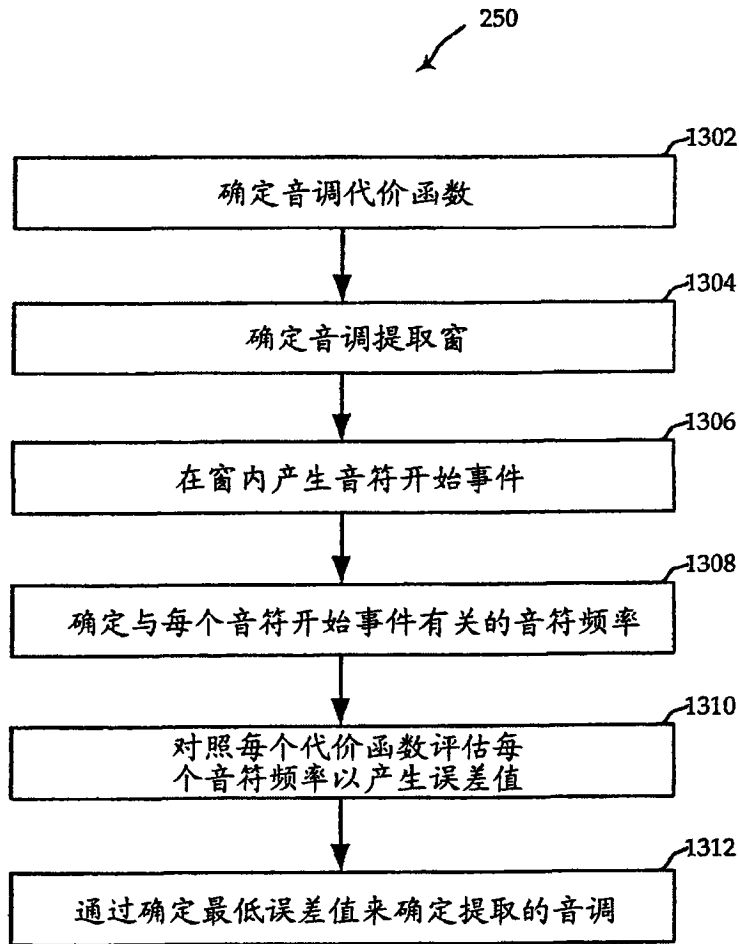


图 13

	C	C#/Db	D	D#/Eb	E	F	F#/Gb	G	G#/Ab	A	A#/Bb	B
1402-1 C 大调	1	0	1	0	1	1	0	1	0	1	0	1
1402-2 C 小调	1	0	1	1	0	1	0	1	1	0	1	0
1402-3 D 大调	0	1	1	0	1	0	1	1	0	1	0	1
A 小调	1	0	1	0	1	1	0	1	0	1	0	1

图 14A

1450

	C	C#/ Db	D	D#/ Eb	E	F	F#/ Gb	G	G#/ Ab	A	A#/ Bb	B
1452-1 C 大调	2	0	1	0	2	1	0	2	0	1	0	1
1452-2 C 小调	2	0	1	2	0	1	0	2	1	0	1	0
1452-3 D 大调	0	2	1	0	1	0	2	1	0	2	0	1
A 小调	2	0	1	0	2	1	0	1	0	2	0	1

图 14B

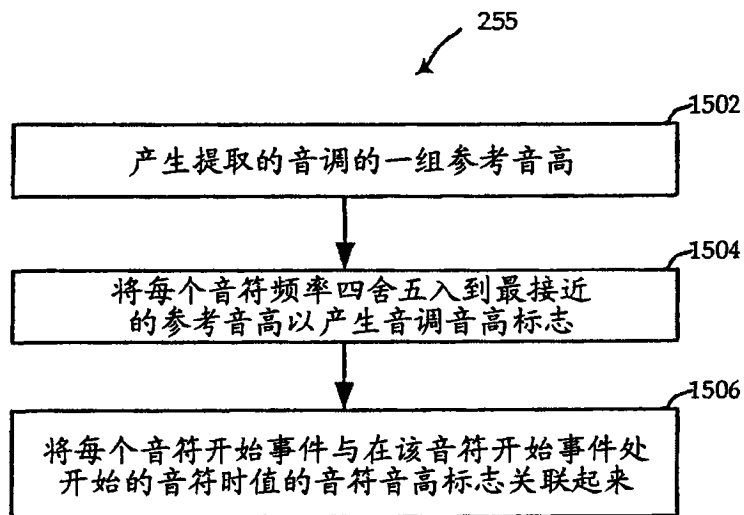


图 15

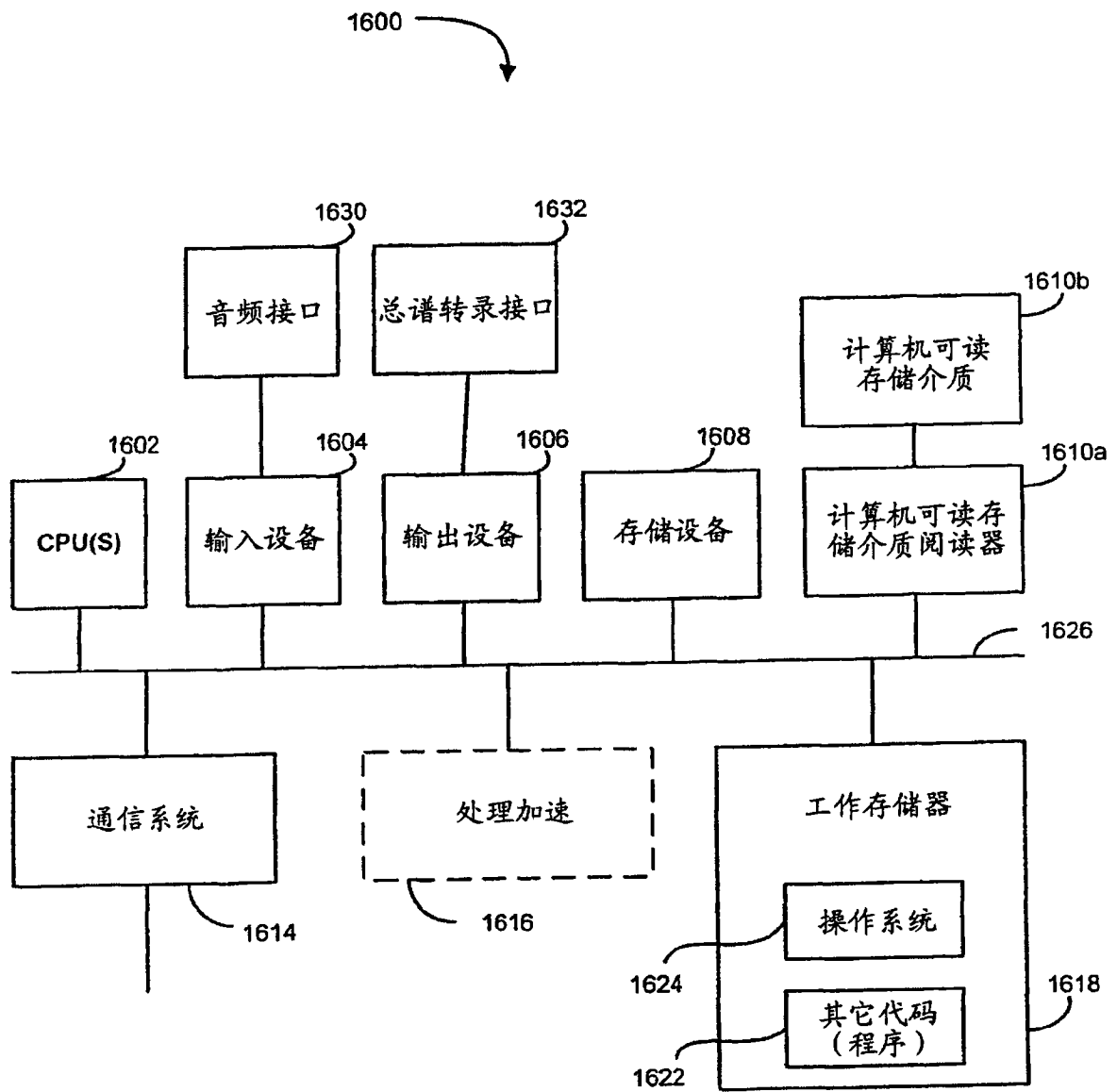


图 16