

专利号 ZL 00818227.2

G10L 21/02 (2006.01)

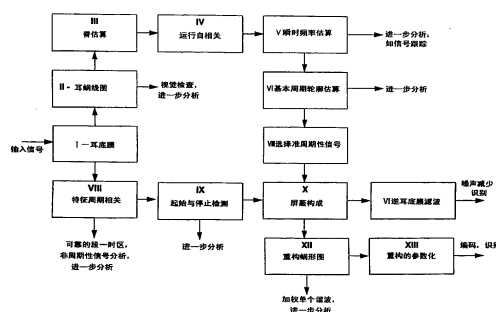
[11] 授权公告号 CN 1286084C

代理人 韩 宏

审查员 杨艳兰

权利要求书 10 页 说明书 56 页 附图 38 页

在步骤1中利用耳底膜模型接收包含目标信号的输入信号。用后续的进一步的步骤，从输入信号滤波目标信号。滤波以后便能将目标信号用作诸如数据压缩的信号识别等进一步处理的输入，也能将目标信号作用在实质上相反的方法上来得到改进的或洁净的信号。



1、一种用于估算输入信号的频率特征的装置，包括：
向其提供信号的耳底膜模型设备，以及
连接在该耳底膜模型设备上的相关器设备，所述相关器设备具有：

连接在该耳底膜模型的一段上，用于接收来源于该段的耳底膜信号的第一输入端，该耳底膜信号出现预定的时段；以及

连接在耳底膜模型设备的同一段上，用于接收移位过可调节的时移 T_1 的耳底膜信号的至少一个第二输入端，及

该相关器设备提供一依赖于时移 T_1 的输出信号，它进一步依赖于存在于该段的耳底膜信号中的频率以及该输出信号构成该信号的频率内容的测度。

2、如权利要求 1 所要求的装置，其中所述可调节的时移 T_1 是调节成对应于该段的特征频率的倒数的。

3、如权利要求 1 或 2 所要求的装置，其中将该耳底膜模型的多个段各连接在独立的相关器设备上，以及其中该装置还包括：

连接在至少若干相关器设备的各自的输出上，用于确定出现在这些段的耳底膜信号中的公共周期的互相关器设备。

4、如权利要求 1 或 2 所要求的装置，其中所述相关器设备是作为配置成用于执行下述数学算法所表示的运算的泄漏自相关器设备实现的：

$$r_{s,T} = L(X_s(t), X_s(t-T))$$

$$S = 1 \cdots S_{\max}$$

$$T = [0, T_{\max}]$$

其中 r 为相关器设备的输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移，及 L 为低通滤波方法。

5、如权利要求 1 或 2 所要求的装置，其中所述相关器设备是作

为配置成执行下述数学算法所表示的运算的低通滤波器设备实现的：

$$r_{s,T} = L(X_s(t), X_s(t-T))$$

$$S = 1 \cdots S_{\max}$$

$$T = [0, T_{\max}]$$

其中 r 为该相关器设备的输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移及 L 为低通滤波方法。

6、如权利要求 1 或 2 所要求的装置，其中所述相关器设备是作为配置成执行用下述数学算法表示的运算的时间归一化相关器设备实现的：

$$R^+_{s,T}(t) = L(X_s(t), X_s(t+T))$$

其中 R^+ 为输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移，及 L 为低通滤波方法。

7、如权利要求 1 或 2 所要求的装置，其中所述相关器设备是作为带有群时延校正器的时间归一化相关器设备实现的。

8、如权利要求 7 中所要求的装置，其中带有群时延校正器的所述时间归一化相关器设备被配置成用于执行用下述数学算法表示的运算的：

$$R^{gd}_{s,T}(t) = L(X_s(t+d_s), X_s(t+d_s+T))$$

其中 R^{gd} 为输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移， d_s 为作为段位置的函数的群时延，及 L 为低通滤波方法。

9、一种用于确定源信号的频谱的装置，包括：

向其提供包含该源信号的输入信号的耳底膜模型设备；及

连接在该耳底膜模型上的相关器设备，

该相关器设备包含：

用于接收来源于耳底膜的一段的耳底膜信号的，用于耳底膜的多个段的各段的第一输入端；以及

用于接收移位过可调节的时移 T2 的耳底膜信号的至少一个第二输入端，及

该相关器设备提供依赖于段的输出信号，该输出信号构成出现在源信号中的能量谱的测度。

10、如权利要求 9 所要求的装置，其中该可调节的时间 T2 设定为依赖于出现在这些段的耳底膜信号中的至少一个公共周期。

11、如权利要求 10 中所要求的装置，其中所述可调节的时间 T2 设定为进一步依赖于耳底膜信号的与段相关的群时延。

12、一种用于确定信号中的峰的装置，包括
向其提供该信号的耳底膜模型设备；

多个积分器设备，各积分器设备与到耳底膜模型的一段的一输入端相连及用于从耳底膜信号中生成激励信号并将该激励信号传递给积分器的输出端；

连接在积分器设备的输出端上，用于确定激励信号中的峰的峰搜索设备；

在该装置中，相关器设备连接在耳底膜模型的段上并且该相关器设备也通信地连接在该峰搜索设备上，并对该相关器设备提供选择的位置上的段的激励信号，该选择的位置及对应地该段依赖于峰搜索设备所确定的峰而及时地变化。

13、如权利要求 12 中所要求的装置，其中所述相关器设备是作为配置成执行用下述数学算法表示的运算的泄漏自相关器设备实现的：

$$R_{s(t), T}(t) = L(X_{s(t)}(t), X_{s(t)}(t-T))$$

其中 r 为相关器设备的输出信号，X 为耳底膜信号，S 为段位置，t 为时间，T 为可调节的时移，及 L 为低通滤波方法。

14、如权利要求 12~13 中任何一项所要求的装置，其中所述相关器设备是作为配置成用于执行用下述数学算法表示的运算的时间

归一化相关器设备实现的：

$$R_{S(t), T}^+(t) = L(X_{S(t)}(t), X_{S(t)}(t+T))$$

其中 R^+ 为相关器设备的输出信号, X 为耳底膜信号, S 为段位置, t 为时间, T 为可调节的时移, 及 L 为低通滤波方法。

15、如权利要求 12~13 中任何一项所要求的装置, 其中所述相关器设备是作为带有群时延校正器的时间归一化相关器实现的。

16、如权利要求 14 中所要求的装置, 其中带有群时延校正器的所述时间归一化相关器是配置成执行用下述数学算法表示的运算的：

$$R_{S(t), T}^{gd}(t) = L(X_{S(t)}(t+d_s), X_{S(t)}(t+d_s+T))$$

其中 R^{gd} 为相关器设备的输出信号, X 为耳底膜信号, S 为段位置, t 为时间, T 为可调节的时移, d_s 为作为段位置的函数的群时延, 及 L 为低通滤波方法。

17、如权利要求 12~13 中任何一项所要求的装置, 还包括：

连接在耳底膜模型设备上用于基本周期轮廓估算的设备,

该设备包含：

连接在耳底膜模型上的输入端；

用于确定估算脊及瞬时周期轮廓的脊确定设备；

连接在脊确定设备上用于选择最高能的平滑瞬时周期轮廓的脊选择器设备；

连接在脊选择器设备上用于将周期轮廓克隆到所有可能的基本周期上的谐波克隆设备, 可能的基本周期的周期轮廓的各组合构成一轮廓假设；

连接在谐波克隆设备上用于选择最高能的周期轮廓的克隆的轮廓选择器设备；

连接在克隆的轮廓选择器设备上用于选择对应于所选择的最高能的平滑瞬时周期轮廓的一部分的至少一种轮廓假设的选择器设备；
以及

连接在选择器设备上用于进一步传递所选择的轮廓假设的输出端。

18、如权利要求 12~13 中任何一项所要求的装置，其中所述信号为声音信号。

19、一种信号识别系统，包括：

输入端；

权利要求 1~18 中任何一项中所要求的装置；

连接在该设备上的存储器，该存储器包含表示要识别的信号的数据；

配置成将来自所述装置的输出的信号与要识别的信号比较并从要识别的信号中确定与所述装置的输出的信号最类似的信号的处理设备；

输出端。

20、如权利要求 19 中所要求的信号识别系统，其中所述要识别的信号表示语音信号。

21、一种数据压缩系统，包括：

输入端；

连接在输入端上的如权利要求 1~18 中任何一项中所要求的装置；

处理器，用于读取所述装置所确定的在该装置的输入端上接收的信号的分量值并将该值传递给

连接在处理器上的输出端。

22、一种数据扩展系统，包括：

用于接收权利要求 21 中所要求的系统所确定的信号分量值的输入端；

用于读取该信号分量值及重构原始信号的处理器装置；

用于输出该原始信号的输出端。

23、一种信号改进系统，包括：

输入端；

第一装置，其是权利要求 1~18 中任何一项中所要求的装置；

用于选择耳底膜信号的部分的屏蔽设备，该屏蔽设备是连接在第一装置的输出上的；

是第一装置的逆反者，用于重构耳底膜信号的选择部分的耳蜗线图，并具有连接在屏蔽设备的输出端上的输入端的第二装置；

连接在第二装置的输出端上的输出端。

24、如权利要求 23 中所要求的信号改进系统，其中所述屏蔽设备包括用于选择相关脊的相关脊估算设备；

用于以正弦响应取代所选择的相关脊的正弦响应加法器设备；

用于如果正弦响应的强度低于原始信号的强度时以原始信号取代正弦响应的加法器设备；

用于消除信号中的不连续性的平滑器设备。

25、一种用于估算源信号的频率特征的方法，包括下述步骤：

在输入端上接收所述源信号；

模拟对具有若干段的耳底膜的所述源信号的响应，从而生成输入信号；

从所述输入信号生成一耳底膜段的至少一个激励信号；

通过用可调节的时移移位至少一个激励信号中的至少一个而生成移位的信号；

将至少一个激励信号中的至少一个与所述移位信号组合，借此得出所述激励信号与至少一个移位信号中该至少一个之间的相关的测度。

26、如权利要求 25 中所要求的方法，其中所述移位信号是从输入信号生成的及该移位信号是与该同一输入信号组合的。

27、如权利要求 26 中所要求的方法，其中所述移位信号是从输

入信号生成的及将该移位信号与用于多个耳底膜段的各个的相同输入信号组合。

28、如权利要求 25~27 中任何一项中所要求的方法，其中所述组合是用包含执行用下述数学算法表示的运算的泄漏自相关步骤执行的：

$$r_{s, \tau} = L(X_s(t), X_s(t-T))$$

$$S = 1 \cdots S_{\max}$$

$$T = [0, T_{\max}]$$

其中 r 为相关器设备的输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移，及 L 为低通滤波方法。

29、如权利要求 25~27 中任何一项所要求的方法，其中所述组合是用包含执行用下述数学算法表示的运算的时间归一化的相关步骤执行的：

$$R^+_{s, \tau}(t) = L(X_s(t), X_s(t+T))$$

其中 R^+ 为输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移，及 L 为低通滤波方法。

30、如权利要求 25~27 中任何一项所要求的方法，其中所述组合是用包含群时延校正步骤的时间归一化相关步骤执行的。

31、如权利要求 30 中所要求的方法，其中所述组合是由包含执行用下述数学算法表示的运算的时间归一化相关步骤执行的：

$$R^{gd}_{s, \tau}(t) = L(X_s(t+d_s), X_s(t+d_s+T))$$

其中 R^{gd} 为输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移， d_s 为作为段位置的函数的群时延，及 L 为低通滤波方法。

32、如权利要求 25~27 中任何一项中所要求的方法，其中将可调节的时间 T_2 设定为依赖于存在这些段的耳底膜信号中的至少一个公共周期。

33、如权利要求 32 中所要求的方法，其中在生成至少一个激励信号的步骤之后，执行用包括所述激励信号的频率分量的所述耳蜗线图的段区确定时间的步骤，以及

所述时移是根据所确定的区调节的。

34、一种确定信号中的峰的方法，包括：

在输入端上接收所述源信号；

模拟对具有若干段的耳底膜的所述源信号的响应，借此生成输入信号；

从所述输入信号中生成一耳底膜段的至少一个激励信号；

确定这些激励信号中的峰；

组合选择的位置的段的激励信号，该选择的位置及对应地该段依赖于所确定的峰而及时地改变。

35、如权利要求 34 中所要求的方法，其中所述组合是用包含执行用下述数学算法表示的运算的泄漏自相关步骤执行的：

$$R_{s(t), \tau}(t) = L(X_{s(t)}(t), X_{s(t)}(t-T))$$

其中 r 为相关器设备的输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移，及 L 为低通滤波方法。

36、如权利要求 34 中所要求的方法，其中所述组合是用包含执行用下述数学算法表示的运算的时间归一化的相关步骤执行的：

$$R_{s(t), \tau}^+(t) = L(X_{s(t)}(t), X_{s(t)}(t+T))$$

其中 R^+ 为相关器设备的输出信号， X 为耳底膜信号， S 为段位置， t 为时间， T 为可调节的时移，及 L 为低通滤波方法。

37、如权利要求 34~36 中任何一项中所要求的方法，其中所述组合是作为带有群时延校正的时间归一化的相关步骤实现的。

38、如权利要求 37 中所要求的方法，其中带有群时延校正的所述时间归一化相关包含执行用下述数学算法表示的运算：

$$R_{s(t), \tau}^{gd}(t) = L(X_{s(t)}(t+d_s), X_{s(t)}(t+d_s+T))$$

其中 R^d 为相关器设备的输出信号, X 为耳底膜信号, S 为段位置, t 为时间, T 为可调节的时移, d_s 为作为段位置的函数的群时延, 及 L 为低通滤波方法。

39、如权利要求 34~36 中所要求的方法, 还包含:

估算基本周期轮廓估算, 包含:

确定估算的脊与瞬时周期轮廓;

选择最高能的平滑瞬时周期轮廓;

将周期轮廓克隆到所有的可能的基本周期上, 可能的基本周期的各周期轮廓组合构成轮廓假设;

选择最高能的周期轮廓;

选择对应所选择的最高能的平滑瞬时周期轮廓的一部分的至少一种轮廓假设; 及

进一步传递所选择的轮廓假设。

40、如权利要求 34~36 中所要求的方法, 其中所述信号为声音信号。

41、一种识别信号的方法, 包含:

权利要求 25~39 中任何一项中所要求的方法;

将输出信号与要识别的信号比较并从要识别的信号中确定最相似于该信号的一个最相似的信号。

42、如权利要求 40 中所要求的方法, 其中所述要识别的信号表示语音信号。

43、一种压缩数据的方法, 包括:

权利要求 25~42 中任何一项中所要求的方法;

读取所接收的信号的用权利要求 25~42 中任何一项所要求的方法确定的信号分量值并进一步传递这些值。

44、一种扩展数据的方法, 包括:

接收用权利要求 43 中所要求的方法确定的信号分量值;

读取这些信号分量值及使用读取的这些信号分量值来重构原始信号；

输出该原始信号。

45、一种改进信号的方法，包括：

第一方法，其是权利要求 25—39 中任何一项中所要求的方法；

选择耳底膜信号的部分；

为第一方法的逆反者，用于利用所选择的部分作为输入重构耳底膜信号的所选择的部分的耳蜗线图的第二方法。

46、如权利要求 45 中所要求的方法，其中所述选择包括：

选择相关的脊；

用正弦响应取代所选择的相关的脊；

如果正弦响应的强度低于原始信号的强度，用原始信号取代正弦响应；

消除信号中的不连续性。

47、如权利要求 45~46 中任何一项中所要求的方法，其中所述声音信号包含来自至少一个说话人的语音。

48、如权利要求 45~46 中所要求的方法，其中所述源信号为未知的信号混合物。

49、如权利要求 45~46 所要求的方法，其中在选择了所述检测的信号步骤之后，进一步分析所述检测的信号。

信号分析方法与装置

技术领域

本发明涉及权利要求1的前序部分中的估算频率特性的装置。再者，本发明涉及权利要求7的前序部分中的估算信号的频谱的装置。本发明还涉及包含这些装置及执行装置的功能的方法的信号识别系统、数据压缩系统及信号改进系统设备。

背景技术

上面所述的装置是从 T. Andringa 在 ESCA EuroSpeech 97, RhodeS, 希腊, ISSN 1018-4074, 2859-2862 页中所公布的“使用听觉模型与泄漏自相关器调谐到语音”已知的。从这一 Andringa 1997 论文中, 已知利用模拟耳底膜的工作原理的听觉模型。将该模型的这些段的输出用作一组调谐到特定的固定周期的泄漏自相关器单元的输入。然后, 作为段号的函数的响应自相关器单元的活动便是要检测的周期性信号的频谱的直接测度。

利用这一已有技术文件中描述的装置, 有可能重构其基本周期是已知且恒定的信号的能量谱。然而, 存在的问题在于当不知道周期, 周期不恒定, 或既不知道又不恒定时, 已知的设备并不提供正确与/或最佳的能量谱重构。本发明的目的为在声音信号的周期不知道与/或不恒定时提供能正确重构该信号的设备。

发明内容

在本发明的第一方面中, 为达此目的本发明提供了一种用于估算输入信号的频率特征的装置, 包括: 向其提供信号的耳底膜模型设备,

以及连接在该耳底膜模型设备上的相关器设备，所述相关器设备具有：连接在该耳底膜模型的一段上，用于接收来源于该段的耳底膜信号的第一输入端，该耳底膜信号出现预定的时段；以及连接在耳底膜模型设备的同一段上，用于接收移位过可调节的时移 T_1 的耳底膜信号的至少一个第二输入端，及该相关器设备提供一依赖于时移 T_1 的输出信号，它进一步依赖于存在于该段的耳底膜信号中的频率以及该输出信号构成该信号的频率内容的测度。

在本发明的第二方面中，为达此目的本发明提供了一种用于确定源信号的频谱的装置，包括：向其提供包含该源信号的输入信号的耳底膜模型设备；及连接在该耳底膜模型上的相关器设备，该相关器设备包含：用于接收来源于耳底膜的一段耳底膜信号的，用于耳底膜的多个段的各段的第一输入端；以及用于接收移位过可调节的时移 T_2 的耳底膜信号的至少一个第二输入端，及该相关器设备提供依赖于段的输出信号，该输出信号构成出现在源信号中的能量谱的测度。

在本发明的第三方面中，为达此目的本发明提供了一种用于确定信号中的峰的装置，包括向其提供该信号的耳底膜模型设备；多个积分器设备，各积分器设备与到耳底膜模型的一段的一输入端相连及用于从耳底膜信号中生成激励信号并将该激励信号传递给积分器的输出端；连接在积分器设备的输出端上，用于确定激励信号中的峰的峰搜索设备；在该装置中，相关器设备连接在耳底膜模型的段上并且该相关器设备也通信地连接在该峰搜索设备上，并对该相关器设备提供选择的位置上的段的激励信号，该选择的位置及对应地该段依赖于峰搜索设备所确定的峰而及时地变化。

在本发明的第四方面中，为达此目的本发明提供了一种信号识别系统，包括：输入端；权利要求 1~18 中任何一项中所要求的装置；连接在该设备上的存储器，该存储器包含表示要识别的信号的数据；配置成将来自所述装置的输出的信号与要识别的信号比较并从要识

别的信号中确定与所述装置的输出的信号最类似的信号的处理器设备；输出端。

在本发明的第五方面中，为达此目的本发明提供了一种数据压缩系统，包括：输入端；连接在输入端上的如权利要求 1~18 中任何一项中所要求的装置；处理器，用于读取所述装置所确定的在该装置的输入端上接收的信号的信号分量值及将该值传递给连接在处理器上的输出端。

在本发明的第六方面中，为达此目的本发明提供了一种数据扩展系统，包括：用于接收权利要求 21 中所要求的系统所确定的信号分量值的输入端；用于读取该信号分量值及重构原始信号的处理器装置；用于输出该原始信号的输出端。

在本发明的第七方面中，为达此目的本发明提供了一种信号改进系统，包括：输入端；第一装置，其是权利要求 1~18 中任何一项中所要求的装置；用于选择耳底膜信号的部分的屏蔽设备，该屏蔽设备是连接在第一装置的输出上的；是第一装置的逆反者，用于重构耳底膜信号的选择部分的耳蜗线图，并具有连接在屏蔽设备的输出端上的输入端的第二装置；连接在第二装置的输出端上的输出端。

在本发明的第八方面中，为达此目的本发明提供了一种用于估算源信号的频率特征的方法，包括下述步骤：在输入端上接收所述源信号；模拟对具有若干段的耳底膜的所述源信号的响应，从而生成输入信号；从所述输入信号生成一耳底膜段的至少一个激励信号；通过用可调节的时移移位至少一个激励信号中的至少一个而生成移位的信号；将至少一个激励信号中的至少一个与所述移位信号组合，借此得出所述激励信号与至少一个移位信号中该至少一个之间的相关的测度。

在本发明的第九方面中，为达此目的本发明提供了一种确定信号中的峰的方法，包括：在输入端上接收所述源信号；模拟对具有若干

段的耳底膜的所述源信号的响应，借此生成输入信号；从所述输入信号中生成一耳底膜段的至少一个激励信号；确定这些激励信号中的峰；组合选择的位置的段的激励信号，该选择的位置及对应地该段依赖于所确定的峰而及时地改变。

在本发明的第十方面中，为达此目的本发明提供了一种识别信号的方法，包含：权利要求 25~39 中任何一项中所要求的方法；将输出信号与要识别的信号比较并从要识别的信号中确定最相似于该信号的一个最相似的信号。

在本发明的第十一方面中，为达此目的本发明提供了一种压缩数据的方法，包括：权利要求 25~42 中任何一项中所要求的方法；读取所接收的信号的用权利要求 25~42 中任何一项所要求的方法确定的信号分量值并进一步传递这些值。

在本发明的第十二方面中，为达此目的本发明提供了一种扩展数据的方法，包括：接收用权利要求 43 中所要求的方法确定的信号分量值；读取这些信号分量值及使用读取的这些信号分量值来重构原始信号；输出该原始信号。

在本发明的第十三方面中，为达此目的本发明提供了一种改进信号的方法，包括：第一方法，其是权利要求 25~39 中任何一项中所要求的方法；选择耳底膜信号的部分；为第一方法的逆反者，用于利用所选择的部分作为输入重构耳底膜信号的所选择的部分的耳蜗线图的第二方法。

在从属权利要求中陈述本发明的特别有利的细节。从下面的描述中呈现本发明的进一步的目的、细节、改型、效果及详情，其中对附图进行了参照。

附图说明

图 1.1 展示了已知耳底膜的图；

- 图 1.2 展示了按照本发明的设备的框图；
- 图 1.3 展示了相对于耳底膜模型的一段确定的自相关函数；
- 图 1.4 展示了相对于耳底膜模型的所有段确定的估算的频率能量谱；以及
- 图 1.5 展示在二维图中合成的图 1.3 与图 1.4 的结果。
- 图 1.6 示出图 1.2 中所示的低通滤波器的第一实例的框图。
- 图 1.7 示出图 1.2 中所示的低通滤波器的第二实例的框图。
- 图 1.8 示出图 1.7 中所示的低通滤波器的第二实例的扩展的框图。
- 图 1.9 示意性示出按照本发明的示例方法的流程图。
- 图 2.1 示出展开的耳底膜的示意性表示及作为时间的函数的 BM 段速度的一些示例曲线。
- 图 2.2 示出字“NUL”的作为时间的函数的 BM 模型的段的运动的二维图或耳蜗线图（cochleogram）。
- 图 2.3 示出在图 2.2 的耳蜗线图中 175ms 上的条的位置的剖面。
- 图 2.4 示出作为段位置的函数的 BM 模型的频率响应的曲线。
- 图 2.5 示出作为不同频率的正弦波的段位置的函数的能量曲线。
- 图 2.6 示出作为段位置的函数的能量分布曲线及能量分布的谐波逼近的选择。
- 图 2.7 示出无噪声信号与加上鸡尾酒会噪声的信号的调谐的自相关（TAC）的曲线。
- 图 2.8 示出从/NUL/中的/U/导出的时间归一化相关图。
- 图 2.9 示出从有与无噪声的信号确定的脊峰的曲线。
- 图 2.10 示出一些脊与从这些脊之一估算的本振频率的自相关曲线。
- 图 2.11 示出从图 2.9 中的/NUL/在 $t=250\text{ms}$ 上估算的自相关曲线。
- 图 2.12 示出作为时间的函数在洁净与噪声条件中估算的瞬时频

率的曲线。

图 2.13 示出句子/NUL EEN TWEE DRIE/的时间的函数的脊与瞬时频率轮廓的耳蜗线图与曲线。

图 2.14 示出图 2.13 的瞬时频率轮廓之逆的本振瞬时周期的曲线及作用在这些本振瞬时周期上的平滑方法的曲线结果。

图 2.15 示出与图 2.14 的平滑的轮廓一致的基波周期轮廓假设曲线及从这些假设选择的基波周期轮廓。

图 2.16 示出从作为时间的函数的图 2.15 的选择的轮廓估算的音调轮廓曲线。

图 2.17 示出根据图 2.16 的估算的周期轮廓的调谐的自相关选择的结果的曲线。

图 2.18 示出一部分时间归一化的相关 (TNC)，图 2.10 的自相关便是从其中导出的。

图 2.19 示出图 2.13 中的 TAC 值大于四分之一本振能量的区的曲线及从这些区中导出的屏蔽。

图 2.20 示出有噪信号、洁净信号及从该有噪信号导出的重新合成信号的图。

图 2.21 示出在重构过程中接连的步骤中的图 2.20 的有噪信号的图。

图 2.22 示出对应于图 2.21 中 $t=275\text{ms}$ 的重构的耳蜗线图的截面。

图 2.23 示出由用于识别的倒频谱 (cepstral) 系数表示的信息的耳蜗线图表示。

图 3.1 示出根据峰值样板在频谱中选择峰值的图。

图 3.2 A—C 示出峰值选择的接连的步骤的结果的图。

图 3.3 A—G 示出输入频率 175Hz、350Hz、525Hz 及 750Hz 的音调分布的曲线。

图 3.4 A—G 示出输入频率 250Hz、375Hz、500Hz 及 750Hz 的音

调分布的曲线。

图 3.5 A—B 示出为图 3.3 的输入频率的组合选择的矩阵与行及得出的最终分布。

图 3.6 示意性示出通用音调估算法的接连的步骤的流程图。

图 3.7 A—B 示出自相关及求和的自相关的曲线。

具体实施方式

下面参照信号中的语音识别描述本发明。虽然本发明能优越地应用在语音识别技术中，本发明适用于通用的信号处理，而并不局限于声音信号或包含语音的信号的处理，此外，虽然是以互相交互作用描述本方法的接连的步骤的，这些步骤同样能应用在不同的方法中，如下面要说明的。

在本申请中，将听觉事件定义为理想地表示单一的源的所有可估算的或可确定的信息的表示。对于许多信号处理方法与设备，尽可能近地靠近这一表示是重要的。

旨在靠近这一表示的大多数已知的（语音）信号处理方法是基于准平稳性假设的。这意味着如振幅与频率内容等信号的某些方面能建模为从能假设为短的时段上（对于语音通常选择大约 10ms 的值）恒定的过程始发的，这一假设的合理性在于信号，即语音，是由不能无穷快地变化的物理系统产生的。

这是广泛地用在本申请中的非常合理的假设。然而，这一假设只对单个源的信号成立。如果信号是由两个源产生的，它将比对单个源所能预期的变化得更快，并且当然更不同。结果，对于单个源有效的准平稳性形式不能应用在混合的源上。在任意的未知环境中，情况甚至更坏，因为可能存在这样的信号成分，对于其准平稳性不再是有用的逼近。如果仍然应用准平稳性，导致的逼近误差将无可挽救地恶化合成信号并因而降低达到信号的正确描述与/或正确识别结果的概

率。

通常通过将信号分成帧并假设接连的帧的序列在整个时间上提供该信号的频率内容的充分足够的描述来实现准平稳性。由于帧的宽度（或窗口的有效宽度）与频率分辨率成反比，引入了时间与频率分辨率之间的折衷。在基于帧的方法中不能最佳地处理在频率细节与时间细节两者上都重要的信号。信息在整个时间上的组合是主要问题。帧的采用引入了不连续性，使得难于确定基础信号分量的连续性。这进一步使得难于将单个源的信号信息赋予单个表示。非矩形窗口的采用及丢弃相位（窗口信号内的时间信息）甚至更恶化这一问题。

这必然导致大量信号处理技术及实际上所有语音信号处理技术不适用于任意信号。具体地，诸如短期傅里叶变换（STFT）、线性预测（LP）及基于帧的滤波器组等技术不得用于任意信号。从而将这些技术应用在这些信号上通常不会成功或者只在非常窄的应用范围内成功。

带有适当抽样时间的准平稳性只能可靠地应用在单个源的信号成分上。对于（声）源的未知混合物，需要更适合的信号处理方式。本发明的目标便是提供这一系统。

因此提供了尽可能好地利用源的规律性的系统。不幸的是由于源尚未分类而源的规律性是未知的。在这一情况中系统只能采取最弱的可能的先有知识。但任何（声）源显示起始、任意的连续发展及停止。结果，不是脉冲样的所有（声）源都具有连续发展的部分。

在语音的情况中，大多数种类的音乐与大量的其它信号，在大多数时间中是明显的连续的发展。在语音的情况中，只有象 /t/、/k/ 或 /p/ 这样的爆破音可能有争论不存在连续的发展。另一方面，象 “Why I owe you an hour?” 这样的发音能以整个发音构成单一的连续整体的方式发音。

频率与振幅的连续性是严格定义的信号特性，它们是大量声源共

享的。这种连续性，假定能从信号中得到证实，便能加以利用而无须关于信号类型的任何进一步知识。因此，信号分量的连续性极适合于协助将单个声源的声音资料赋予听觉事件。只要信号分量呈现连续的发展，它便很可能发源单一的源。这是相当可靠的结论，因为不相关的源导致互相平滑地符合的信号分量的概率较小。此外，诸如基本频率轮廓等信号特性也是连续的，并有助于将不同信号分量分组在一起：单个准周期性声源的所有谐波保持基本频率的整数倍。与某一本频率轮廓符合的频率轮廓很可能属于同一源，或者如音乐中通常的情况，将源乘以一相关时间发展。

结果，信号分量的连续性构成将单个源的信息赋予单个表示的最可靠的提示之一。然而这一过程并不完整，时间与频率上的连续性应保持。

图 1.9 中示出了按照满足上述限制的本发明的将单个源的信息赋予单个表示的示例方法的流程图。

首先在步骤 I 中从耳底膜模型设备生成耳底膜的段的 BM 信号。这一 BM 信号是基于作用在耳底膜的输入上的信号的。在标题为“BM 模型”开始的节中将更详细地描述这一步骤。

第二，在步骤 II 中积分该 BM 信号以便为各段的激励提供测度。根据这一积分能提供段 $\times$ 时间 $\times$ 激励测度的一个三维矩阵。这一矩阵的可见表示称作耳蜗线图。在标题为“耳蜗线图”开始的节中详细描述这一步骤。

第三，在步骤 III 中根据三维矩阵提供的信息，在该矩阵中估算称作脊的单个信息分量。这些单个分量为能估算出瞬时频率轮廓的段-时区，在标题为“脊估算”开始的节中将更详细地描述这一步骤。

第四，执行脊下的自相关，这在步骤 IV 中为轮廓的周期性提供一种测度。在标题为“在脊下运行自相关”开始的节中将更详细地描述这一步骤。

第五,分析这些自相关值以便在步骤 V 中为基本周期轮廓的发展的各时刻确定本振频率。这一本振频率在段与时间中是本地 (local) 的。在标题为“本振瞬时频率轮廓”开始的节中将更详细地描述这一步骤。

第六,在步骤 VI 中利用本振频率估算信号的音调。虽然正式地说名词音调与基本频率是不能互换的,前者指称(可物理地测定的)后者的(主观)感觉,但整个本文个中将可互换地使用这两个名词。在标题为“基本周期轮廓估算”开始的节中将更详细地描述这一步骤。

第七,在步骤 VII 中应用将 BM 模型的段的激励关联的方法。由于不相关的激励不太可能属于同一信号,通过取所有相关的信号分量,便逼近了该信号。在标题为“调谐的自相关”开始节中将更详细地描述这一步骤。

第八,在步骤 VII 中应用确定哪些耳底膜段呈现对应于该段本身的特征频率的周期性的方法。这是用激励与按照该段最佳周期移位后的激励之间的相关执行的。这一步骤中所确定的信息能用于估算周期性与诸如噪声及起始等非周期性信号分布。在标题为“特征周期相关”开始的节中将更详细地描述这一步骤。

第九,在步骤 IX 中应用起始检测设备。这一步骤通过搜索特征周期相关设备的输出或耳蜗线图上的急剧变化检测起始与停止。在标题为“起始检测”开始的节中将更详细地描述这一步骤。

第十,在步骤 X 中应用屏蔽构成方法。这一屏蔽设备利用这一阶段得到的信息来标记可能从目标源发源的时间与地点的区。可定义各标记不同特性的区的多个屏蔽。在标题为“屏蔽构成”开始的节中将更详细地描述这一步骤。

第十一,在步骤 XI 中可应用逆耳底膜滤波设备。这一步骤利用屏蔽设备所标识的区逆滤波耳底膜的激励来达到表示屏蔽下的信号信息的重新合成的信号。在标题为“逆耳底膜滤波”开始的节中将更

详细地描述这一步骤。

第十二，步骤 XII 中可应用耳蜗线图重构设备。这一步骤利用从信号估算的并由屏蔽表示的信息来测定单个信号分量并通过将这些单个信号分量相加来重构信号。在标题为“耳蜗线图重构设备”开始的节中将更详细地描述这一步骤。

第十三，在步骤 XIII 中，可将参数化设备应用在重构的耳蜗线图上。这一步骤提供适合于作为识别设备的输入的信息。在标题为“参数化”开始的节中将更详细地描述这一步骤。

BM 模型

按照本发明的方法从生成 BM 信号开始。在听觉系统中，从声音，即压力波动，到神经信息的转换是在称作耳底膜的结构周围进行的，如图 2.1 中示意性地所示。耳底膜是能用传递线路物理学描述的相关物理结构。传递线路是在时间与空间上都连续的结构，其中在耳底膜的情况中空间对应于频率，因为耳底膜上各位置具有其本身的特征频率。结果，耳底膜将声音振动转换成神经信息以便保持时间与频率（通过其与位置的对应性）的连续性供进一步处理。

图 2.1 示出耳底膜 1 的主要特征的非常示意性表示。耳底膜为一位于耳蜗中的长度为 3.5cm 的螺旋结构 2，耳蜗为大约 1cm^3 的蜗牛壳样结构。靠近蜗牛壳开口处的耳底膜侧对大约 20KHz 的频率最敏感，越往耳蜗里面对各位置最敏感的频率按照（近似的）对数的位置频率关系下降到 20Hz，耳底膜的频率范围因此是三个数量级或大约 10 个 8 度音阶。大约 3000 个纤毛细胞均匀地沿耳底膜分布，将局部振动转换成分级电位，后者又被编码成动作电位并由 30000 个神经元传递给脑干。这些神经元的轴突构成听觉神经。

在所示的示例性设备中，采用了耳底膜的一维传递线模型，如从 Duijfhuis, H., HoogStraten, H.W., van Netten, S.M., Diependaal,

R.J., 及 Bialek, W. (1985) 已知。“用耦合的 Van der pol 振荡器模拟耳蜗隔膜”, 周围听觉机制, 编辑 J.B.Allen, J.L.Hall, A.E.Hubbard, S.T.Neely 与 A.TubiS (Springer, 纽约) 290—297 页。该模型最有关的特性为时间与位置的连续性及一对一的位置-频率关系。这要求能将耳底膜模型解释为带有物理地耦合的滤波器的滤波器组: 相邻的滤波器在所有的时间点上呈现类似的位移。然而, 虽然耳底膜 (BM) 模型类似于哺乳动物听觉系统的部分, 本发明不局限于具有哺乳动物听觉系统的这种相似性的传递线模型。为了清楚起见, 在本申请中将名词 ‘BM 模型’ 用于具有耳蜗的类似功能的任何非均匀的传递线。

有可能应用不同的耳底膜模型, 例如象实际耳底膜那样非线性的已知耳底膜模型。在本发明的这一示例实施例中采用了该模型的有利的线性型式。虽然通常认为非线性模型具有更好的性能, 由于它更紧密地类似于人的听觉系统, 但发现能以令人满意的性能实现线性模型。BM 模型设备的这一线性型式只需较少的处理并且设计与实现较为简单。

这一线性能够作为重叠与相加滤波器组的高效实现, 并有助于解决中心问题: 如何分离信号的混合物。总之线性引起可加性, 可将其解释为能够分裂信号 a 与 b 的混合物而不引入依赖于 a 与 b 两者的截项。在大多数非线性系统中不能保证的截项的不存在, 简化了信号分离系统的设计与实现并且只需较少的处理。

此外, 已知的耳底膜模型具有 400KHz 的内部更新频率并包括跨越整个人类频率范围的 400 段。为了进一步减少示例实施例中的处理时间, 所示的 BM 模型是作为带有跨越 30 与 6100Hz 之间的频率范围的 100 个频道的滤波器组实现的。这一滤波器组实现需要 20KHz 的输入与输出样本频率。借此与已知的模型相比, 获得计算效率的提高。

图 2.1 中, 示出了内耳的耳底膜的一部分电等效电路图。图的左

侧，能提供信号 VSound，它由多达 400 个互相耦合的一系列二级滤波器（称作段）处理。各二级滤波器代表一定的质量惯性（线圈）、劲度与阻尼，它确定所提供的信号 VSound 中的振动被互相连接的二级滤波器减弱得多快。建模在实际内耳上，耦合发生在将声音信号感觉传递到脑的神经束的各个段上。

所描述的 BM 模型有利地应用在上面参照图 1.9 所描述的方法中。然而，BM 模型能有利地应用在任何信号识别系统中，只要该系统利用 BM 信号。

耳蜗线图

诸如基于 FFT 的能量谱图等时间—频率图认为是表示解释信号的最有关信息的。不幸的是，它在时间与频率两者上都是不连续的。通过平均各耳底膜段的（重叠的）帧的能量能计算出在位置上（并间接地在频率上）连续的频谱图样时间—频率图。然而这一过程蕴含应当避免的准平稳性，因为输入尚未识别为保持准平稳性的信号。

这些问题是通过在 FFT 频谱图的时间与位置（频率）两者中都采用连续的替换元来解决的。这一替换元为位移的泄漏积分平方或耳底膜段的速度的平方。速度（位移的一阶导数）的使用优于位移的使用，因为速度的使用增强了高频分量，这降低了低频分量对高频分量的屏蔽效应。泄漏积分描述系统在各时间点上丢失关于其前一状态的信息但知道关于当前状态的信息的过程。在本实施例中选择下述泄漏积分：

$$r_s(t) = r_s(t - \Delta t) e^{-\Delta t / \tau} + x_s(t) x_s(t), \quad s = 1 \cdots s_{\max} \quad (1)$$

这一公式中 $r_s(t)$ 表示时间 t 上段 s 的泄漏积分能量的值， Δt 为抽样周期， $t - \Delta t$ 表示上一次抽样的时间， $x_s(t)$ 为频道的当前输出值。这一一阶系统的时间常数 τ 确定存储器范围。对于大的 τ 值，指数非常接近 1，对于小值，指数的影响变得更明显，因为它降低了

$r_s(t)$ 的前一值的作用。平方项 $x_s(t) x_s(t)$ 为非负的。因此 $r_s(t)$ 也是非负的。

可将式 1 推广成：

$$r_s(t) = L(x_s(t)) \quad (2)$$

在该等式中函数 L 表示任何形式的低通滤波。因此可应用任何种类的低通滤波，然而应用泄漏积分器特别有利，因为泄漏积分器所执行的功能与人类听觉系统的功能相似，这在下面要进一步说明。信号 $x(t)$ 可以是本征激活的测度的任何形式。因此可利用半波整流速度、位移或加速度、或给出本征激活的（非负）测度的任何其它形式。然而采用速度平方特别有利，因为假设速度是人类听觉系统中的纤毛细胞的驱动力，而这一值的平方计算是简单的运算，它给出能量测度。

低通滤波之后可执行其它步骤来确定信号。例如可应用耳蜗线图的简单视觉检验。为了提供在自动化方式中获得的信号的可靠逼近，必须提供进一步的步骤。本发明提供这些步骤，如下面将说明的。然而，本发明不限于这些特定的进一步的步骤。

在整个本申请中将 τ 值保持在 10ms 上，但能加以修改。实际神经元也执行泄漏积分过程并且对于听觉系统中的神经元 10ms 是正常值。虽然式 1 的输入为耳底膜速度的平方，神经生理学等价物是全正幅值压缩的半波整流的耳底膜速度。半波整流是由 Corti（皮盾）器官中的纤毛细胞执行的。自然系统呈现通常作为立方根逼近（ $\sim x^{0.3}$ ）BM 运动 x 的动态范围压缩。

动态范围压缩对于将所有相关特征集中到同一范围内是必要的。这是重要的因为按照式 2 计算的 $r_s(t)$ 由于诸如语音等自然信号的特性而具有能达到 50dB（分贝）或以上的动态范围。为了补偿式 2 中的平方，加倍立方根的效应并用 $x^{0.15}$ 逼近：

$$R_s(t) = [r_s(t)]^{0.15} \quad (3)$$

这一非线性动态范围压缩专用于视觉表示。除非明确地指明，所

有视觉表示都将采用这一形式的压缩。

由于泄漏积分过程是低通滤波过程，可将输出 $r_s(t)$ 下抽样到积分时间常数的数量级上的抽样率。为了提供锐利的起始，选用了 200Hz 的抽样率，这对应于每 5ms 一个样本。这导致耳蜗线图成为所要求的双连续时间-频率图。图 2.2 示出女性说话人说出的荷兰语字 /NUL/ (英文: ZERO) 的耳蜗线图。此字是目标句子 /NUL EEN TWEE DRIE/ 的一部分，在整个本申请中作为未知信号的示例使用这一句子。虽然样本是声音信号，本发明并不限于语音或声音，如前面所述。

在图 2.2 中，在大约 $t=50\text{ms}$ 与 $f=220\text{Hz}$ 开始的宽带是对应于基频 f_0 的一次谐波 h_1 。在发音到 350Hz 以上的值期间基频上升。一次谐波上面并与之平行的带是二次谐波 h_2 。最低的少数谐波构成第一共振峰 F_1 。在 $t=120\text{ms}$ 上从 /N/ 过渡到 /U/ 之后可见到第二共振峰 F_2 ，并在 /L/ 期间从 2000Hz 下降到几百 Hz。注意共振峰位置的这一改变导致不同的谐波作为最明显的局部频率成分互相接续。在 /N/ 期间勉强可见第三共振峰 F_3 ，但在其余的发音期间成为明显的。在更高的频率区中，可见第四甚至可能第五共振峰。

从 /U/ 到 /L/ 的过渡是平滑的，由于从鼻音 /N/ 到元音 /U/ 的过渡而导致从 /N/ 到 /U/ 的过渡是部分地不连续的。在 /N/ 结束时舌尖离开硬腭，除了鼻腔以外使口腔也起作用。注意该字的起始不连续性是锐利的但停止是平滑的。这是由于泄漏积分过程的指数衰变及耳底膜与式 3 的非线性的结合的呼出 (ringing-out) 效应。

图 2.3 中描绘了 $t=175\text{ms}$ 上的耳蜗线图的垂直截面。该图示出对应于图 2.2 中的垂直线下的信息的作为段号 (上方水平轴) 或对应的频率 (下方轴) 的函数的能量分布图。注意尖峰结构。在较低的段号上尖峰对应于分辨出的谐波。对于较高的段号，单个谐波分辨得不太好并最终合并成共振峰。这一表现是非线性位置-频率关系的直接结果。图中描绘了若干谐波。前三个、第九、十三、十八及二十五谐波

支配响应。第四至八谐波刚能分解，至于第十至十二谐波只存在最小的可见证据。这些谐波是被其它分量（部分地）屏蔽掉的。虽然更高的谐波不能分辨，它们对共振峰的形状仍起作用并对元音/U/的音色起作用。

段的引开(entrainment)是传递线模型的非常重要特性，并且是由耳底膜构成单一连续结构这一事实引起的。当一个明显的信号分量驱动某一段时，该段将一起拖动其相邻的段而它们又一起拖动它们相邻的段等。这一效应作为位置的函数迅速衰减。只有能克服其它信号分量的补充效应的信号分量会达到局部支配及产生尖峰。引开和屏蔽一样在高频侧上比低频侧上更突出。

大多数图采用已根据实验估算的位置-频率关系。在原始 BM 模型中未耦合的段的共振频率是按照 Greenwood 位置-频率关系选择的，其中 X 是从峰尖以 mm 测定的：

$$f_c = 190 \cdot 10^{0.6 \times [mm]} - 145 [\text{Hz}]$$

$$S = \frac{400}{3} \cdot \frac{x[mm]}{3.5[mm]} \quad (4)$$

段下标 S 可通过使用 BM 的长度、原始模型中的段数及本实施例中实际上只使用三分之一的段这一事实与 x 相关。

段之间的交互作用将特征频率向稍低的值移位。图 2.4 以虚线示出 Greenwood 位置-频率关系。用细黑线描绘作为段号的函数的真实特征频率。它永远在 Greenwood 位置-频率关系下方。与段的特征频率相似，每一频率具有一特征段。实际位置-频率关系是基于测定各段对单位振幅的对数间隔开的正弦曲线的范围的响应强度的。其值为对数编码的灰度级得出的矩阵构成图 2.4 的背景。注意，高频段对低频刺激的响应比低频段对高频刺激的响应大。

由于（准）周期性信号包含正弦波成分的组合，研究 BM 对单一频率刺激的响应是有用的。BM 对一定频率的响应称作正弦响应。图 2.5 中描绘了若干实例。与驱动频率无关，所有 BM 响应都具有类似

的不对称形式，带有向高频侧比向低频侧更明显的尾部。

图 2.5 描绘只能在足够次数（诸如 10 次）振动与/或少数（诸如 5）次积分时间常数 t 之后才能达到的稳定状态情况。自然信号很少呈现改变得慢到足以全面证明这一稳定状态假设的信号分量。在 BM 的低频侧上，在达到稳定状态所需的 50ms 或以上的期间，音调与振幅很少足够恒定。这得出比理想的正弦响应宽的响应。在 BM 的高频侧上，较快达到稳定状态，但自然信号的音调中的随机波动同样加宽响应。

积分器与对应的耳蜗线图可用在上面参照图 1 描述的方法中，然而在振荡信号分量需要激活测度的所有情况中，都能使用上述过程。

估算单个信号分量

由于本信号识别系统中使用的 BM 模型是线性的，它的响应是对驱动声源的单个分量的响应之和。在准周期性声源 $S(t)$ 的情况中，可将输入描述为：

$$s(t) = \sum_n a_n(t) h_n(t), \quad h_n(t) = \sin\left(\frac{2\pi n}{T(t)} t - \phi_n(t)\right) \quad (5)$$

其中 $a_n(t)$ 为谐波成分 $h_n(t)$ 的振幅。谐波为周期轮廓 $T(t)$ 及相位函数 $\phi_n(t)$ 的函数。

这一信号的耳蜗线图是用式 1 定义的。除了这里不予考虑的某些例外现象之外，平方与足够长的积分时间常数 t 保证相位项 $\phi_n(t)$ 的作用消失。在大多数情况中，与式 1 的泄漏积分过程的时间常数 t 的值相比， $a(t)$ 改变得较慢。这意味着 $a_n(t)$ 能作为因子 $\langle a_n^2(t) \rangle$ 确定 $h_n(t)$ 的耳蜗线图成分的比例常数对待（短时间间隔）。 $\langle \rangle$ 表示泄漏积分过程估算的时间平均值。 $h_n(t)$ 的耳蜗线图成分表示为 $R[h_n(t)]$ 。对于缓慢地发展的 $h_n(t)$ ， $R[h_n(t)]$ 能用图 2.4 与图 2.5 中给出的正弦响应逼近。

这意味着按照式 1 从信号 $S(t)$ 得出的耳蜗线图能近似地表示为：

$$R(t) = \sum_n a_n^2(t) R[h_n(t)] \approx \sum_n w_n(t) R_n(t) \quad (6)$$

$R_n(t)$ 为用对应于本振瞬时频率 $h_n(t)$ 的时间发展的最佳段的正弦响应的依次逼近的单位振幅谐波成分 $h_n(t)$ 的响应。权重 $w_n(t)$ 确定这一正弦响应的定标。

原则上，正弦响应的权重 $w_n(t)$ 是未知的并应从信号中估算。为了估算图 2.3 中的信号的单个谐波的成分，发现两种不同的方法是适用的。

本申请人开发的第一种方法通过忽略向低频侧的屏蔽来利用正弦响应中的不对称性。在这一情况中，通过首先加权对应于基频的频率的正弦响应来逼近图 2.3 中的信号。这计及二次谐波位置上的一部分激励，认为余数是二次谐波引起的。在下一谐波的位置上，减去前一谐波的成分并认为余数是当前谐波引起的。能将这一过程继续到谐波频率超过最后的段的特征频率为止，但实践中它限于分辨子谐波的 BM 区。因此，这一方法对一次谐波工作得特别好并可用于频谱重构，如下面说明的。

本申请人开发的第二方法是找出矩阵方程 $Rw=E$ 的数值解。在这一方程中， E 为目标耳蜗线图截面， R 为与单个谐波的频率关联的正弦响应的集合而 w 为所要求的加权值。当应用在图 2.3 中的信号上时，基频 f_0 为 $1/4.60=217\text{Hz}$ ，如能从图 2.8 中的 TNC 估算的。相关的谐波频率为 nf_0 。BM 的最后一段的特征频率为 6100Hz ；因此能表示的最高谐波号为 28。对于各频率，可选择一正弦响应并将其加在矩阵 R 上。解 $W=R^{-1}E$ （在最小平方意义上）并将 w_n 的负值设定为零而得出图 2.6 中的结果。

图 2.6 上方用连续线描绘目标 E ，下方表示各谐波的定标的成分 w_n 。上方的点划曲线给出正弦响应的加权的和。如能看出的，匹配得非常好，而前三个共振峰的谐波成分能可靠地估算。最高谐波的权重只能在共振峰附近可靠的估算。在其它位置上与谐波关联的正弦响应

几乎完全重叠而数值误差可能影响结果。较低的基频加重这一问题，但使用 BM 模型中较多的段缓解这一问题。

当基频轮廓已知时，这一第二方法是用于分析周期性信号的谐波内容的高效与相当精巧的方法。当信号的音调急剧改变时这一技术也能奏效。在这一情况中，必须通过选择反映谐波的本振瞬时频率的频率集合来计及群时延效应的作用。然而，如果提供了正确的音调轮廓，这一校正是简单的。

调谐的自相关

为了分裂其信号源不确定的信号混合物，需要使用最弱的可能的基本假设，即使用最一般的信号特性。重要的一般特性是信号是否准周期性的。在语音、音乐及许多其它信号中，准周期性信号表示时间与能量的最大组成部分。注意完全周期性信号不经常出现。大多数自然信号呈现由源特性引起的幅度与/或频率调制。这限制了 Andringa 1997 中提出的已知系统的有用性。

如果对于各谐波 $y_i(t)$:

$$y_i(t) = y_i(t + T(t)) \quad (7)$$

则将声源 $y(t)$ 定义为带有基本周期轮廓的准周期性的。

如果声源的谐波 $y_i(t)$ 带有耳底膜的段 s ，该段的响应 $x_s(t)$ 将同样呈现准周期性。结果：

$$x_s(t) = x_s(t + T(t)) \quad (8)$$

如果 $T(t)$ 是已知的，式 8 能与式 2 组合而产生：

$$\begin{aligned} r_{s,0}(t) &= L(x_s(t), x_s(t)) \\ &= L(x_s(t + T(t)), x_s(t)) \\ &= r_{s,T(t)}(t) \end{aligned} \quad (9)$$

这意味着在 $T(t)$ 是正确的基本周期轮廓的条件下， $r_{s,T(t)}(t)$ 紧密地逼近声源 $y(t)$ 所补充的所有段的耳蜗线图成分。这是重要的

因为 $T(t)$ 是具有对声源 $y(t)$ 独一无二的非常高的概率的信号特性。将值的集合 $r_{s, T(t)}(T)$ 定义为调谐的自相关 (TAC), 因为它基于自相关值并调谐到基本周期轮廓 $T(t)$ 上的 (并从而也到基频轮廓 $f_0(t) = 1/T(t)$ 上)。

式 9 只对正确的基本周期轮廓成立。对于与目标源的轮廓不相关的基本周期轮廓, $x_s(t)$ 与 $x_s(t+T)$ 的值不相关而它们的平均值将接近零。这意味着 TAC 对于正确地估算的周期轮廓具有类似于耳蜗线图的能量测度的值而对于随机选择的或不相关的周期轮廓具有接近 0 的值。

$$r_{s, T(t)}(t) = \begin{cases} r_{s, o}(t) & | T(t) \text{ 是正确的} \\ 0 & | T(t) \text{ 是不正确的} \end{cases} \quad (10)$$

这一特性构成将特定声源的信息赋予听觉事件的基础。

当不知道准周期性源补充了哪些段时, 用下式计算所有的段的 TAC:

$$\begin{aligned} r_{s, T(t)}(t) &= L(x_s(t), x_s(t+T_s(t))) \\ &= L(x_s(t), x_s(t+T(t+d_s))) \end{aligned} \quad (11)$$

这一公式包含对与段有关的群时延的校正。线性系统的响应时间能形式化成群时延。可将线性系统的群时延定义为段 s 的平方脉冲响应 $h_s(t)$ 的重心:

$$d_s = \frac{\sum_t t[h_s(t)]^2}{\sum_t [h_s(t)]^2} \quad (12)$$

群时延导致对应于该周期轮廓的谐波的表示中的与段相关的时移。图 2.7 中的结果是用正确的与段相关的周期轮廓 $T_s(t) = T(t+d_s)$ 计算的:

$$\begin{aligned} r_{s, T(t)}(t) &= r_{s, T(t)}(T - \Delta t)e^{-\Delta t/\tau} + x_s(t)x_s(t+T_s(t)) \\ s &= 1, \Lambda, s_{\max} \end{aligned} \quad (13)$$

图 2.7 上方示出字/NUL/（图 2.2）的耳蜗线图及相关 TAC 的正值。下方示出在加上导致 0dB 的信噪比（信号与噪声能量相等）的鸡尾酒会噪声时这一信号的耳蜗线图。右下方示出相关的 TAC。与上面的图比较，保持了大多数主导结构。TAC 不是在整个 500ms 上定义的，因为只有在存在声源时才定义声源的周期轮廓。注意将 TAC 图的负值只在可见的图中设定为零。在整个本申请中都将如此。

从正确地估算的周期轮廓中得出的调谐的自相关表示与这一轮廓一致的准周期性信息。不保证所有信息都属于同一个源，然而保证表示引开 BM 区的目标源的所有周期性成分。

调谐的自相关是非常坚韧的。这具有若干理由。首先，调谐的自相关选择受目标谐波支配的所有段范围。在诸如语音等宽带信号的情况下，其中少数谐波或共振峰起支配作用，得出尖峰的耳蜗线图。其它声音的共振峰或类似结构产生甚至更明显地成功支配相同区的甚至更强的峰的概率通常是小的（但不是零）。这一概率当然极大依赖于信噪比（SNR）及能量在频率范围上的分布。以在 -6dB（比=1:4）的信噪比上屏蔽目标语音的普通宽带信号，目标语音的未屏蔽的峰的数目减少到难于为搜索听觉事件找到一组可靠的起点的水平。在这些条件中人类语音理解力急剧恶化。

TAC 的坚韧性的第二理由在于源不需要处于支配地位来提供一致的局部成分。只要较不明显的源的平均成分 $X(t)X(t+T)$ 大于局部地处于支配地位的源的平均值 $X'(t)X'(t+T)$ ，该较不明显的源即使是光学地屏蔽的也将提供正的成分。由于不存在峰，这一情况并不为听觉事件估算提供可靠的起点。这可以说明为什么在听众不知道预期何时在第一次提出时不能理解某些有噪声的句子，而在听众能形成正确的预期时，相同的句子是可以识别的。例如，天真的听众对于 -6dB 的 SNR 上的目标句子可能有困难，而有经验的听众能理解 -10dB 或以下的目标句子。

TAC 的应用的最重要的问题是正确估计基本周期轮廓 $T(t)$ 的必要性。由于它不是可以直接得到的，必须从信号中估计它。存在着许多音调估算技术，但其中没有一种在任意（有噪声的）信号上能正确地执行。调谐的自相关能与坚韧的音调估算技术结合有利地应用，它在任意有噪声信号上正确地执行。在关于音调估算的节中提供这一技术。

现在首先参见图 1.1，其中示出内耳的耳底膜的电工学等效电路图的一部分。图中左侧，能提供输入信号 V_{Sound} ，它受到一系列多达 400 个互相耦合的称作段的二阶滤波器的处理。各二阶滤波器代表一定的质量惯性（线圈）、刚性与阻尼，它确定所提供的信号 V_{Sound} 中的振动被互相连接的二阶滤波器衰减得多快。模拟实际的内耳，耦合代替到达传送声音信号印象到大脑的神经束的不同的段的位置。

在图 1.1 中所描绘的耳底膜的电气等效电路模型的辅助下，按照本发明的设备是可以工作的。为达到这一目的，将图 1.1 的耳底膜模型包含在图 1.2 中的具有参照数字 1 的框中。对该耳底膜模型作用一可能受噪声污染的声音信号 2。对耳底膜模型 1 的一或多段连接一低通滤波器 3。这一低通滤波器 3 产生构成输入信号 2 的估算的输出信号 4。

有利地选择包含在低通滤波器 3 中的滤波器函数。为了达到这一目的，在本发明的第一变型中，将该低通滤波器设计成具有第一与第二输入的乘法器，同时对第一输入作用一起源于耳底膜的一段信号并将其作用预定的时间段，对第二输入上推后可调节的时间 $T1$ 作用该信号，而该乘法器提供一依赖于时间 $T1$ 的输出信号，后者与实质上出现在该段的信号中频率相关并构成声音信号的频率内容的测度。

在本发明的另一变型中，将低通滤波器设计成具有第一与第二输入的乘法器，而对于耳底膜的各段，对第一输入作用起源于该段的信号，推后时间 $T2$ 将该信号作用在第二输入上，而乘法器提供一与段

相关的输出信号，它构成在该时间段中实质上出现在声音信号中的频率能量谱的测度。

在更精确的公式表示与图形表示中，能进一步阐明本发明如下。

为了计算频率内容及频率能量谱，利用图 1.6 中的框图，其中 $x_S(t)$ 表示来源于耳底膜模型 1 的信号，而 T 表示可调节的时移。

这里 $\alpha = e^{-\delta t/T}$ ，并且（为 t 与 T ）进一步利用时轴，它是定标成使得 $\delta t = 1$ 的。

在用于确定频率内容的设备的实现中，来自耳底膜模型 1 的信息信号是一个段序列 $S(t)$ ，这是一分段连续函数，因此在图 1.7 中的图形表示中的预定的时间段内，段号是不变的。输出集合 $\{r_{S(t)}, \dots, r_{S(t), T_m}(t)\}$ 表示在时间 t 上的段轮廓下的当前自相关。

图 1.3 示出在特定的时间 t 上述输出集合的典型曲线，其中 T 指示在 X 轴上。

在用于确定频率能量谱的设备的实现中，从耳底膜模型 1 的段的测定结果表示低频的出现相对于与较高频率相关的段带有一定延时。对于给定的信号 $T(t)$ ，这需要用公式 $T_S(t) = T(t + d_S)$ 加以校正，其中 d_S 为段 s 的局部群时延。

这导致图 1.8 中所示的下述实现。

在这一情况中为耳底膜模型 1 的各段确定能量内容，得出与时间 t 相关的输出集合 $\{r_{1, T_1(t)}, \dots, r_{sm, T_{sm}}(t)\}$ 。

图 1.4 用图形表示对应的测定结果，将耳底膜模型 1 的段号标在 X 轴上。

进一步参照表示自相关值的图形表示及作为变量在 x 轴上表示时移 T 及在 y 轴上表示耳底膜模型的段号的图 1.5。这里要指出用于确定一段的频率内容的设备的实施例给出对应于水平截面的测定结果，诸如能为具有大约 470Hz 的共振频率的耳底膜模型的段确定的。用于确定耳底膜模型的所有段的频率能量谱的设备给出对应于垂直

方向上的截面的测定结果，例如在 $t=4.7\text{ms}$ 上所取的截面。

时间归一化的相关图

可将式 13 推广到更通用的连续自相关函数：

$$r_{s,T}(T) = L(x_s(t), x_s(t+T)) \quad s=1, \dots, s_{\max} \\ T \in [0, T_{\max}] \quad (14)$$

这一函数通常作为维度（段数） $\times$ （时段数）的时间展开的矩阵实现。在本申请中称作时间归一化的相关图（TNC）。可以用类似方式将 TNC 的正值描绘成 TAC 频谱图。它示出在图 2.8 中。

这一图示出在 NUL 的/U/的中央 $t=175\text{ms}$ 的 TNC。 $T=0$ 上的垂直线对应于图 2.3 中所描绘的能量谱。 $T=4.6\text{ms}$ 上的垂直带表示基本周期 T_0 的 TAC。这一带对于 $2T_0$ 被重复大约 9.2ms 。这些带构成随着单个谐波的频率增加而变窄的大型垂直结构的峰。各宽带准周期性源具有类似的结构，其总体结果是由瞬时基本周期专门确定的。

关键重要的是 TNC 是在时间、周期性及位置（位置与频率相关）连续的。名称 TNC 是从下述事实得出的：其在式 14 中的定义保证如果一准周期性信号在时间 t_0 上开始，其 TNC 便在这一时间上开始建立而与该信号的周期 T 无关：对于 $t < t_0$ ， $x(t) \times x(t+T(t))$ 的时间平均值接近零，而在 t_0 之后它是大且正的并与 $T(t)$ 的值无关。这一起始时间归一化形式有助于研究所有类型的源的时间发展。在本节中稍后给出比较相关图的不同定义的更细致的讨论。

由于不相关的源不太可能呈现类似的瞬时基频的发展，不同的源的垂直结构重叠的概率是低的，在所有声源互相重叠表达的 $T=0$ 上能量项的情况不是这样。引入周期性作为附加的信号维度不仅允许分裂周期性与非周期信号的混合物，并且也允许分裂准周期性信号的混合物。注意这是部分地理想化的：二或多个准周期性信号的组合导致单个 TNC 的叠加，这是比单一 TNC 更难于解释的。

TNC 的垂直截面对应于所有段 s 的自相关滞后 T 。水平截面对应于单一段的全运行自相关。对于非周期性信号，相关性将作为 T 的函数急剧下降，但由于这一源是周期性的，自相关具有余弦的外观。注意大多数段是受单一谐波控制的。这对于对应于较低谐波的段最为明显。局部运行的自相关的周期性反映作为时间的函数的段的主驱动力的频率。再现在所有段中的第一周期为对应于 217Hz 的 4.60ms 。对于二次谐波，第二周期在 4.6ms 上达到峰值。这如预期的那样对应于 $1/(4.6/2) = 434\text{Hz}$ 的瞬时频率。刚超过 2000Hz ，BM 的区受九次谐波控制。这一区对应于第二共振峰。注意由于受九次谐波屏蔽，十次谐波的位置不能估算出。第三共振峰紧接在 3000Hz 下面得到表示，并受 2826Hz 上的十三次谐波控制。注意 TNC 能以非常高的精度确定瞬时本振频率。这是避免基于帧的方法及保持连续性的直接结果。下面进一步说明本振频率估算算法的细节。

TNC 是极为丰富的表示，但其最重要的特征在于 TNC 能表示通过时间 (t)、位置 (S) 与周期性 (T) 的任意连续路径。

这意味着如果我们知道或假设一种周期轮廓 $T(t)$ 作为源特性，我们便能研究 $T(t)$ 作为时间的连续函数的后果。反之，如果已知段序列 $S(t)$ 表示单一声源的信息，便有可能利用 TNC 来研究由段序列 $S(t)$ 下的运行自相关性表示的信息的发展。使用沿段的依赖于时间的路径是有利的，因为对于大量各式各样的信号，信号的基本周期在时间上不是恒定的，诸如语音声与音乐声。

注意 TNC 使用相关性方向的精心选择。对于基于帧的自相关根据：

$$r(n) = \sum_{t=0}^N x(t)x(t \pm n) \quad x(t) = 0 \text{ 如果 } t \notin \{0, 1, \dots, N\} \quad (15)$$

这使得相关性是否基于对应于 $t+n$ 或 $t-n$ 的滞后没有什么差别。只要将所有成分相加，结果是一样的。惯例是选择减号，因为这类类似于没有时延的因果系统。对于连续地更新（或运行）的自相关，选择

‘t-n’导致与基于‘t+n’的实现不同的信息时间排序。

本节中将讨论基于泄漏积分的相关图的三种不同的连续实现。然而，本发明能应用于许多其它的相关器变型。第一种选择为：

$$r_{s,T}^{-}(t) = L(x_s(t), x_s(t-T)) \quad s=1, \dots, s_{\max} \\ T \in [0, T_{\max}] \quad (16)$$

这一等式是时间展开的矩阵的矩阵元素的定义函数。符合式 11，表示段 s 的自相关值与自相关滞后 T 。上面标记的“-”指称最后项中的减号。 s 维度的矩阵下标跨度，及 T 维度的下标跨度。再一次， $x_s(t)$ 表示 BM 段 s 的输出。这一实现是因果的，无须时延。

第二种实现，在时间上向前看定义 TNC：

$$r_{s,T}^{+}(t) = L(x_s(t), x_s(t+T)) \quad (17)$$

丢掉下标，将这一实现称作 $r^{+}(t)$ 。因为未来的信息在得到它以前不能处理，这一实现带来当前时间 t 与能得到整个相关图的时刻的时间 $T_{\max}$ （诸如 12.5ms）之间的时间差。

通常执行一种形式的群时延归一化。在归一化期间，将群时延作为延时对待，而不是作为延时与能量表达式的时间模糊（smearing）的组合。群时延归一化是通过用局部群时延 d_s 时移耳底膜段的响应来执行的。将时间 t 上的高频段的信息与 30ms 以后或在时间上更晚的低频段信息组合。这导致有时方便的能量表达式的表示。群时延归一化可用‘+’或‘-’号执行。这里只考虑‘+’型式，但‘-’型式同样可能：

$$r_{s,T}^{gd}(t) = L(x_s(t+d_s), x_s(t+d_s+T)) \quad (18)$$

与式 14 比较甚至需要更长的延时。虽然式 14 中的最小延时为 $T_{\max}$ ，现在的延时为 $T_{\max} + d_s$ 。 $T_{\max}$ 的倒数是在相关图中要表示的最低频率。对于语音，有用的下限为 80Hz 或 12.5ms，而对应的群时延为 28ms。组合延时最终大于 40ms。比这一延时更重要的可能是 $r^{gd}(t)$ 表示在时间上原始相隔 30ms 或以上的信息。以 5ms 的帧抽样率，

这对应于 6 帧或以上。

从保持连续性的观点，式 18 保持时间连续性不好于其它实现。在这些实现中相邻的段 a 与 b 之间在时间 t 上的改变包含最小的时间间隔，而在群时延归一化的情况中，有效的改变为：

$$S_a(t) \rightarrow S_b(t + \Delta t + d(S_b) - d(S_a)) \quad (19)$$

因为在相邻的段之间群时延稍有不同，引入在信号的急剧改变期间有可能降低连续性假设的有效性的附加时移。由于 TNC 的连续性是本假设（即在处理期间不能检验的假设），违背连续性可能导致不能预测的结果。然而，群时延本身是这里使用的传递线模型中的位置的连续函数。这意味着在与群时延的组合中连续性仍然是有保证的。

自然信号很少是稳定的，因此必须选择最适合于反映不稳定信号的相关图型式。这便是 TNC，因为与 $r^-(t)$ 及 $r^{gd}(t)$ 比较，它反映起始及它们出现时的信号改变。再者 $r^-(t)$ 的建立方式导致这一综述的其余部分的任务与解决方法的上下文中所有种类的实际问题；特别是在起始与急剧改变期间。因此 $r^-(t)$ 不是适当的选择。减号与群时延归一化的组合进一步使这一点复杂化。这便是不首先考虑它的原因。

使用群时延归一化不是必要的。然而，与 $r^-(t)$ 不同， $r^{gd}(t)$ 可以使用而无附加的问题。它提供带有部分地纠正了的群时延效应的可见表示。

脊的估算

用 TNC 获得的瞬时局部频率信息构成未知的噪声环境中的音调轮廓的最佳估算的基础。计算上 TNC 是极度低效的，因为它具有数量级（段数） $\times$ （每秒样本数） $\times$ （周期数）。对于 100 段、20KHz 的抽样频率及 25ms 的最大周期（500 个不同值），这对应于每秒 10^9

× (2 次乘法 + 1 次加法)。虽然有可能可观地提高计算效率，但需要更高效方法。

这一方法是通过在耳蜗线图中确定有可能提供关于单个信号分量（例如谐波）的明显信息的区发现的。如在图 2.3 的上下文中所讨论的，各信号成分试图补充耳底膜的一个区。这意味着各位置通常是受单一的信号分量控制的。此外，所有这些区在对应于该信号分量的频率的位置上显示峰值。这意味着各峰对应于单一的信号分量。诸如图 2.3 中十次至十二次谐波等信号成分几乎完全被更强的成分屏蔽而并不显示出独立的峰。当将搜索空间缩小到耳蜗线图上的峰时，便高效地选择能可靠地估算单个信号分量的信息的位置。

为了减少虚假的峰的数目，可通过透过时间将峰组合成脊。丢弃不能分类为可靠的脊（其选择在关于可靠性测度的节中更详细地说明）的成分的所有峰位置。这得出示出两个互补的图的图 2.9：左方给出在叠加在/NUL/的耳蜗线图上的 0dB 鸡尾酒会噪声中估算的脊。右方示出补充信息：带有在洁净的信号中估算的脊的有噪声的耳蜗线图。

这些图允许在噪声中估算的脊对在洁净环境中估算的脊之间进行方便的比较。在有噪声的信号中估算的脊通常与洁净目标的最明显的峰重合。由于这些脊是从有噪声的信号估算的，它们也表示源信息控制的位置。如在右侧图中所见，鸡尾酒会背景主要包含其它说话人的语音中的强度峰。由于这些强度峰比整个声源持续较短，这些脊较不可靠。对于包含大量不相关的源的背景，或包含非周期性噪声的背景，通常是这种情况。这一要求有助于解决问题，由于通过精确定位最重要的信息源便高效地缩小了搜索空间。

当两个信号成分具有对应于单一的段或其近邻的频率时出现重要的情况。在这一情况中，带有相长与相消的干扰的周期交替出现。这导致以两个信号分量之间的频率差的倒数的周期的调幅，并在对应

于两个分量的加权平均频率的位置上形成脊。与这一脊关联的泄漏积分能量值显示调幅。在有噪声的情况中，这可能得出中断的脊。

在信号分量对应于相隔很远的段时，以使两者都携带它们对应的最佳段，在信号分量之间产生另一重要的交互作用。段之间的某处（由于屏蔽的不对称性特性通常选择高频段），存在着感受来自两个分量的相当的影响的段。这些段必须跟随两种不同频率而不破坏 BM。结果平均局部振幅（并从而对应的能量）是小的。对应的局部能量同样小。这不可避免地导致带有用山谷分开的两个峰的情况。对于持续一定时间的信号成分，对应的峰串一起构成时间的脊。这证明存在着对应于分隔得充分远的连续发展的信号分量的稳定的脊。

运行自相关的估算

本振瞬时频率轮廓是用运行自相关估算的。采用新方法来确定脊，现在具有了一组连续的脊 $\{S_i(t)\}$ 并且由于 TNC 是在时间 t 与位置 S 上连续的，便有可能如下计算沿脊 $S(t)$ 的运行自相关：

$$r_{S(t), T}(t) = L(x_{S(t)}(t), x_{S(t)}(t+T))$$

$$T=[0, T_{\max}] \quad (20)$$

随着峰位置平滑地改变，其关联的自相关也如此。注意与式 13 的调谐的自相关的对称性。该等式表示在具有作为时间的函数的周期轮廓 $T(t)$ 的所有段上的一组函数，而式 20 为带有作为时间的函数的段序列 $S(t)$ 的所有 T 上的一组函数。TAC 描述 TNC 的垂直截面而运行自相关描述水平截面。

这样估算的自相关的典型实例给出在图 2.10 左方。上方自相关是作为大多数脊的代表的典型的良好构成的实例。在一些情况中，自相关呈现频率成分的混合物。下方自相关给出其中的三次谐波部分地被二次谐波屏蔽的实例。这导致混淆的自相关及更复杂或甚至不可能的本振瞬时频率（LIF）估算（如在下一节中描述的）。本系统假设良

好地构成的自相关。不良构成的自相关偶而出现并导致并不对应于段特征频率的（不正确的）LIF 估算。发现可将这些值丢弃。

本振瞬时频率轮廓（LIF）的估算

图 2.10 上方的自相关对应于时间 $t=285\text{ms}$ 上的图 2.9 的目标信号的四次谐波。选择这一情况是因为它并不对应于非常明显的脊且本振频率改变得相当快。本振瞬时频率能通过计算平均峰间距离来逼近，在本例中为 12.0ms 中的 13 次振动。这相当于 1083Hz 。但是如将要示出的，这是得出 $t=285\text{ms}$ 与 $t=285+12\text{ms}$ 之间的平均本振频率的准平稳性的不必要的应用。

使用峰间距离的发展的一阶逼近改进了本振瞬时频率估算。这能通过取峰间样本中的距离并通过这些值装配一阶模型来实现。 $T=0$ 上的自相关峰的这一模型的值产生 LIF 的估算。这是高效的方法，但这一方法遭受时间离散化效应，因为与 0.94ms 的本振瞬时周期相比， 0.05ms 的抽样周期不是无足轻重的小的。

时间离散化效应可通过用三点二次配合改进峰位置的估算来减小。这得出峰位置估算的十倍改进。右方图示出对应于重新估算的峰间距离的频率与通过这些值的线性配合。在峰 0 的位置上得出的 LIF 值为 $1064 \pm 5\text{Hz}$ 。误差小于 0.5% 。对于语音，音调的波动阻碍较高的精度。注意 LIF 每 10ms 改变 25Hz 或 2.35% 。每 10ms 2.35% 的改变率对应于每秒因子 10，或 3.3 个八度。这些是自然语音的自然值。

因为避免了固定的时间窗口，本振瞬时频率估算是在时间与频率两者上都非常精确的。采用开窗口的信号的本振瞬时频率估算受到准平稳性应用及时间与频率分辨率之间的折衷的限制：

$$\Delta t = \frac{1}{\Delta f} \quad (21)$$

窗口宽度的选择确定时间分辨率以及频率分辨率。增加窗口的尺寸降低时间分辨率 Δt 但提高频率分辨率 Δf ，反之亦然。

诸如假设信号由单一谐波复合体构成等关于信号的附加假设能提高瞬时频率估算，但对于任意信号这些假设的有效性没有保证。

基于 TNC 的 LIF 估算同样服从式 21，但是以不同的方式。它是根据引开的并限于导致脊的信号成分。这意味着不能分辨带有对应于相邻的段的频率的两个分量，因为它们导致单一的脊。对于一个分量屏蔽其它分量的情况也是这样。这样增加 BM 段的数目及削尖响应曲线能改善这一无能。但是按照式 21 削尖响应曲线会导致增加群时延。在传递线中，可将式 21 中 Δt 解释为群时延，而 Δf 表示调谐曲线的尖锐性的测度。

因此，基于 TNC 的 LIF 估算的精度限制在产生脊的信号分量上。它进一步受峰的数目及在运行自相关中的峰位置估算的精度限制。并最终受一阶逼近的有效性的限制。在缓慢地改变的信号成分的情况下，形成了稳定的脊并能将自相关的最大滞后选择为表示允许非常精确的 LIF 估算而不使一阶逼近无效的若干周期。注意与一阶配合关联的误差提供这一估算的可靠性的测度。例如，在音调估算期间可利用这一测度。

图 2.11 给出在时间 $t=250\text{ms}$ （见图 2.9）上从有噪声的/NUL/估算的少数自相关的实例。下方示出图 2.9 的左侧图中在 $t=250\text{ms}$ 上的脊 2、4、6、7 与 8（从最低的脊开始编号）的运行自相关。这些自相关暗示这些脊从属于同一源的谐波产生：它们都符合 4.10ms （ 244Hz ）的周期性。上方示出不符合这一周期性脊的 1、3 与 5 的自相关。其中脊 3 与 5 在 2.9 、 5.8 或 8.7ms 的周期性上可能符合。

本振瞬时频率估算是对噪声坚韧的。在图 2.12 中计算与描绘两种条件下的 LIF 值：这些点是从洁净的/NUL/估算的本振瞬时频率之值。这些星号是从有噪声的/NUL/估算的。注意洁净信号中的大多数频率成分保持清楚地出现在有噪声的环境中。更接近的检验显示波动通常小于百分之二。这表示脊为单个信号分量的估算构成非常可靠的

信息源。

基本周期轮廓估算

可靠而坚韧的音调估算技术的研制是困难的。其主要原因在于识别这些源以前通常不可能确定哪些信号成分或信号特性属于某一源。这是不能确定信号类型的直接后果，例如不能识别该信号便不知道该信号是否为语音。

虽然这一问题通常是不可解决的，但仍可利用诸如平滑地发展的谐波等某些特征。只要用户确保不出现其它类型的声音便能利用它们，不幸的是这不是人们对能应付未知情况的系统的要求。

研制了两种基本周期轮廓估算技术：一种是本节中描述的用于洁净信号的：洁净信号的音调估算，而另一种是下一节中描述的用于噪声污染的信号或包含来自多个源的信号的信号。

有噪声的信号的坚韧的基本周期轮廓估算

本节描述了估算复杂信号的周期轮廓的技术的实施例。它分成两部分。第一部分给出概述，第二部分提供更详细的阐述。

这一特性用于图 2.14 下方所描绘的轮廓。图 2.15 上方示出与图 2.14 的平滑后的轮廓相一致的所有基本周期轮廓假设。一些基本周期轮廓假设互相平滑地重叠或延伸。这是对周期轮廓来源于同一源的强有力的暗示：不相关的周期轮廓构成相容的整体的概率是小的。下面描绘根据三条主要标准的上方图的选择：轮廓必须具有一定的最小长度，例如 50ms，它们必须充分平滑及在多个并发的轮廓的情况中只选择最长的轮廓。这导致强烈的缩减并且通常得出包含一个或更多地正确的音调轮廓候选者的集合。

最后的步骤将剩下的并发候选者与图 2.14 中所描绘的原始局部

周期性信息进行比较来确定哪一个候选者说明最多的周期值，并且为了防止八度误差而具有合理的奇次与偶次谐波比率。最好地符合这些要求的候选者构成算法的最终输出。

图 2.16 示出从带有不同信噪比的混串噪声的信号估算的音调轮廓之间的比较。除了在起始与停止期间的一些差别，该算法能为-3dB 与更好的 SNR 找出正确的轮廓。当算法产生正确的轮廓时，匹配通常是很好地在实际值的 1% 以内。该算法识别出目标声音的最明显的谐波仍相当地能在这些条件中局部地控制的区及进一步利用周期性信息来找出组合尽可能多的这些区的音调轮廓。因为这些区中的周期性信息仍然是实质上未受损害的，音调轮廓必定具有与在洁净条件中估算的类似的质量。在起始与停止期间局部信噪比更不利使得更难以明白地确定周期轮廓。

由于音调轮廓估算技术寻找长的、平滑的及很好地保持的基频轮廓，它寻找能成立的证据的所有组合。

噪声中的基本周期轮廓估算

可靠而坚韧的音调估算技术的研制不是简单的，因为在识别信号之前不可能确定哪些信号成分或信号特性属于目标类型。为诸如语音等在复杂、未知与可变的环境中产生的信号研制了第一种周期轮廓估算设备。它被设计的很坚韧。本节中将给出这一设备的描述。为无噪声信号研制了第二种基本周期轮廓估算设备：其功能依赖于目标不受噪声污染这一假设。这一设备是基于周期性源的所有谐波呈现公共的周期性这一特性的。本节中将提出这一发明：无噪声信号的基本周期估算。

发现可靠的信息源的标记为：

- 1、各时刻上最有活力的脊
- 2、长脊

3、平滑的脊

4、带有对应于本振特征频率的频率的脊

其中组合了多数这些特征的脊特别可靠。提出的实施例采用在完整的发音上工作的方法，但有可能以提供仅有 100ms 或更少的延时的周期轮廓假设的估算的方式重新实现该设备。这一延时的下限是由群时延效应、计算本振频率所需的时间范围及最重要的允许系统产生的周期假设的数目的组合确定的。具有 50ms 延时的系统比允许它在 100ms 上综合信息的系统拥有较少可利用来减少好像是基本周期候选者的数目的信息。可供选择地，延时取决于信号本身：非常可靠的信息需要小的延时，而较不可靠的信息需要较多及更长的处理。

本节描述依赖于在大多数环境中工作很好的一组固定标准的设备。下一节提出能取代的一实施例的固定标准来改进实现的灵活性与可靠性的一组可靠性测度。

本实施例采用周期域，其它实现可采用频域。

图 3.6 提供坚韧的基本周期轮廓估算算法的五个步骤的概述。

步骤 1 估算脊与本振瞬时周期轮廓（LIF 轮廓的倒数）。步骤 2 选择最可靠的瞬时周期轮廓。步骤 3 将这些轮廓克隆成所有可能的基本周期轮廓。步骤 4 将克隆的周期轮廓组合成可靠的基本周期轮廓假设。步骤 5 是供选用的并选择最可靠的基本周期轮廓，以便在各时刻上只存在单个基本周期轮廓假设。

算法的输入为图 2.9 中所表示的信息并包括耳蜗线图、脊及本振瞬时周期。

第一步骤已在关于 LIF 估算的节中描述过。

算法的第二阶段为最有活力的脊的选择与平滑。这一算法从检测其对应的最佳段与脊的最佳段相差一段以上的瞬时周期开始。用段特征周期（也称作最佳周期）取代这些周期值。可获得的信息描绘在图 2.14 上方。为了看起来清楚只描绘了小于 5ms 的周期（即频率高于

200Hz)。由于这一表示是基于 0dB 的信噪比的情况的，它示出了必须丢弃的大量虚假成分。

跟随各脊，并且只要接连的周期互相在 5% 以内便将它们赋予同一周期轮廓。当两个接连的周期不在 5% 以内时，执行附加的检验来检验下一个值是否在 5% 以内。如果能找到有效的下一个值，使用其相邻的值平均值填补这一间隙，否则开始一个新的轮廓。用轮廓的平滑后的型式 $p(t)$ 加大所有轮廓（见图 2.14 的下方图）。平滑是用 5 点（25ms）线性逼近执行的。在轮廓中心平滑的局部周期是基于每边两帧的局部邻域的。在轮廓的两个第一或最后的点上，平滑后的周期值是基于二阶逼近的对应值的。最后，计算各轮廓的平均顺序。该顺序是用能量表示的相对重要性的测度。最具活力的脊的段具有顺序 1，次最具活力的段具有顺序 2，等等。每当一个周期轮廓的长度超过 50ms 且其平均顺序小于或等于 2 时，或作为替代每当其长度超过 75ms 时，便接纳该周期轮廓。

平滑脊 $p(t)$ 可以也可以不来源于目标语音的谐波。假设各轮廓从头到尾表示单个谐波号。如果该谐波号 n 是已知的，则基本周期 $p_0(t)$ 是已知，因为：

$$p_0(t) = p(t)n \quad \text{或} \quad f_0(t) = \frac{f(t)}{n} \quad (22)$$

作为进一步的限制，有效基本周期值限于 2.5ms（400Hz）与 13.3ms（75Hz）之间的值，这是大多数说话人跨越的范围。例如周期 $p=6\text{ms}$ 可以是基本周期 $p_0=12\text{ms}$ 的二次谐波或 $p_0=6\text{ms}$ 的基波的结果。周期 $p'=2\text{ms}$ 可表示 2 至 6 范围内的任何谐波号。它对应于集合 $\{4, 6, 8, 10, 12\}\text{ms}$ 中的任何 p_0 。如果 p 与 p' 来源于同一源，它们共用相同的基本周期 p_0 。在本例中，6 或 12ms 之一。

如在图 2.15 上方所见，将平滑后的周期轮廓乘以各可能的谐波号并克隆到所有可能的基本周期上。这包含改变这些轮廓的对应特征段，并由于各段具有其本身的群时延，这蕴含按照下式的时移：

$$t \rightarrow t + d(S_{np}) - d(S_p) \quad n \in \{1, 2, \dots\} \quad (23)$$

$d(S_p)$ 与 $d(S_{np})$ 为与分别对周期 p 及周期 np 最敏感的段关联的群时延。注意这一时移蕴含地将瞬时基本周期定义为基波的周期。

第三步骤将克隆的轮廓组合到平滑的基本周期轮廓假设中。这是复杂的过程，因为通常能以不同方法组合轮廓。当两个克隆的轮廓的本振周期平均地落入彼此的 3% 以内时，便将它们组合到单一的假设中。当能估算出在 25ms 期间两个轮廓在 3% 内匹配的二次配合时，便将互相部分地或完整地延伸的轮廓组合。式 22 的时移非常重要因为它能在多个轮廓之间进行可靠的比较。当不存在这一形式的群时延校正时，在音调的急剧改变期间，不能组合同一源的轮廓。最后，丢弃短于 50ms 的基本周期假设。这导致强烈的缩减，并且通常得出包含或多或少地正确的音调轮廓候选者的集合。这一集合描绘在图 2.15 的下方。基本周期轮廓假设的当前集合可构成适当的输出。但对于在每一时间间隔中最多选择一个周期轮廓是重要的应用，有可能选用最后的步骤。

这一供选用的第五与最后步骤包含在并发的轮廓假设之间的强制的选择。当应用在标准的 ASP 系统中时，这是非常重要的阶段，因为它确定将要构成哪些听觉事件。这一选择决定哪一部分信号将按照识别系统的期望及限制来解释。这一阶段的错误将导致识别错误。这保证根据所有可得到的信息（即所有脊及它们对应的瞬时周期）的非常仔细的决策过程。这一决策过程为每时刻选择最多单一的周期轮廓。这一选择的假设结合偶次与奇次谐波的完整分布使得能要求作为可能的谐波的瞬时周期值的数目最大。

基本周期轮廓假设 $p(t)$ 要求的谐波数是通过计数满足下式的瞬进周期值的数目确定的：

$$\cos\left(\frac{p(t+ds)}{p_{s,t}} 2\pi\right) > 0.95 \quad (24)$$

$p_{s,t}$ 是在时间 t 上从段 s 的脊导出的瞬时周期值。而 $p(t+d_s)$ 为用值 d_s 纠正了群时延来表示在段 s 的位置上预期的瞬时基本周期的基本周期假设。再一次群时延校正是必要的，因为比较了耳底膜的不同区的瞬时频率信息。上述公式的条件等价于接受预期的值附近 5.1% 的偏移。

这一公式的变型可用于计数在预期的值 5.1% 以内的奇次与偶次谐波的数目：

$$\begin{aligned} N_{p(t)} &= N_{p(t)}^0 + N_{p(t)}^e \\ &= \sum_i [\cos(\frac{p(t+d_i)}{p_i} \pi) < -0.95] + \sum_i [\cos(\frac{p(t+d_i)}{p_i} \pi) > 0.95] \end{aligned} \quad (25)$$

下标 i 指称所有可能的周期值 $p_{s,t}$ ，而 $p(t+d_i)$ 为在段 s 中时间 t 上所反映的局部瞬时基本周期所需的群时延校正值。 $N_{p(t)}$ 为接受的谐波的总数， $N_{p(t)}^0$ 与 $N_{p(t)}^e$ 为奇次与偶次谐波数。在本例中，奇谐波落在最小值附近，而偶谐波与余弦函数的最大值重合。方括号表示布尔值：如果命题为真则为 1，如果命题为伪则为 0。

两种或以上并发假设的最佳假设便是使下式最大的一种假设：

每帧平均谐波数 · 奇谐波分数 (fraction)

$$= \frac{N_{p(t)}}{L} \cdot \frac{N_{p(t)}^0}{N_{p(t)}} = \frac{N_{p(t)}^0}{L} \quad (26)$$

L 为用帧数表示的基本周期轮廓假设的长度。两种标准都是重要的。要求的谐波的平均数为假设的质量的测度：要求每帧大量谐波的短假设通常比要求每帧少数谐波的较长假设优先选用。当基本周期轮廓是太低的八度时奇谐波的分数是低的。这相当经常发生，因为虚假的成分趋向于增加要求的谐波的平均数。组合的标准减少每帧的奇谐波的平均数。这一简单的标准具有高概率来选择正确的假设。

用所有要求的谐波重新估算及平滑所选择的假设。平滑类似于第一阶段中所应用的平滑。由于所有要求的谐波都对估算有所帮助，数据点的数目大得多。各帧中的局部线性逼近现在是基于在群时延校正

之后对应于 25ms 的环境的所有数据点的。当存在对应于较小环境的 25 个数据点时，局部估算便是基于这 25 个点的。利用式 22 将基本周期轮廓定义为基波的瞬时周期轮廓。为了符合周期性信号成分的选择一节中的选择算法的要求，时移周期轮廓来反映源的瞬时周期而不是基波的瞬时周期。

该算法的最终输出为定义各帧上的一阶逼近的一序列参数。一个参数给出局部瞬时基本周期，另一个给出时间导数。由于两者都是在时间上平滑过的，实际周期可在估算值附近波动。在周期性信号成分的选择中描述的技术决定最终与最优值。

图 2.16 示出从具有混串噪声的不同信噪比的信号估算的音调轮廓之间的比较。除外起始与停止期间的某些差别，该算法能为 -3dB 与更好的 SNR 找出正确的轮廓。当算法生成正确的轮廓时，匹配通常是严格地在实际值 1% 以内的。该设备识别目标声音的大多数明显的谐波仍相当能够局部地控制的区并且该设备利用周期性信息来找出组合尽可能多的这些区的音调轮廓。因为在这些区中的周期性信息实际上是未受到破坏的，音调轮廓必定具有与洁净条件中估算的类似的质量。在起始与停止期间局部信噪比更不利，使得难于明白地确定周期轮廓。

实验表明在 95% 以上的情况或者在具有高于 0dB 的 SNR 的大多数有噪声的环境中更多的情况中，该算法能估算出正确的周期轮廓。在这些情况中它能进行非常好的 TAC 估算。在 0 与 -3dB 之间，正确估算的概率降低到 70%，在 -3dB 以下由于固定的参数而设备不可靠。利用下一节中提出的可靠性测度能提高可靠性。

在下一节中提出另一方法来确定时间中的各（抽样）时刻上（步骤 3）的可能音调值。然后能以脊的估算中所描述的过程（如在关于脊估算的节中所描述的）的类似方法连接这些值来构成音调轮廓。在下一节中还概述了这些音调轮廓的总体可靠性评分的构成。

这一替代方法能以最小的延时确定音调轮廓，如在本节中前面所提到的。

可靠性测度

如上所述，基本周期轮廓估算选择最可靠的信息。当必须在系统（选择器设备）的最高级别上确定可靠性测度时，在较低级别上确定的可靠性测度扮演重要的角色。某一级别上的可靠性测度是从较低级别获得的以及在该级别本身获得的新信息的可靠性测度构成的。从最低到最高级别可定义下述可靠性测度：

1) 在时间中一定时刻上输出的能量中的单个峰的可靠性。这一测度取决于局部信噪比（SNR）。

2) 脊的（整体）可靠性可从为用于构成该脊的峰所确定的（局部）可靠性值构成。其它影响可来自脊的长度与脊的连续性（在位置与能量两者中）。

3) 频率轮廓的可靠性取决于其下方确定频率的脊的可靠性。频率的连续性也可起作用，以及自相关与标准形状的配合也起作用。

4) 在选择脊时，不确定可靠性测度但通过只选择能假设为属于同一源的脊而对较高级别上的可靠性起作用。

5) 音调轮廓的（整体）可靠性是从下述各项构成的：

- 对确定音调/脊与这一音调轮廓配合起作用的脊的可靠性
- 在音调确定方法内确定的配合良好性测度
- 音调的连续性

6) 最终选择的可靠性测度可基于该选择所根据的音调轮廓的可靠性。并且能确定与利用可能出现在每一个位置上的屏蔽量来给出取决于时间与位置的可靠性评分。

上述提到的测度将更详细地说明如下。参见图 3.1，下面更详细地说明峰搜索与局部 SNR 估算。在每一时刻上，在 BM 模型的能量

输出中能找到峰。这些峰是（作为位置的函数的能量的）局部最大值。使用二次内插可用比该模型中所使用的离散化步骤更高的精度确定峰的位置。下一步骤便是搜索具有最靠近局部最大值的峰的峰样板（在预定的峰样板组中）。这是为所有的峰单个地进行的。

下面，必须确定与样板（grSpec）配合得很好的能量谱（ES）的点。一种采用 ES 与 grSpec 之间的垂直距离的配合方法在样板的陡坡上导致不良配合，而在较浅的区上要好多。因此构成一种方法，它利用将 ES 的点放在样板 grSpec 上所需的旋转角。这一方法从峰位置上开始。它首先检验这一位置左边的点。如果能找到 ES 上的点与 grSpec 上的点的方向相距不大于 α 度，但假设能将该 ES 点放在样板 grSpec 上，并将注意力转移到下一点上。只要将 ES 点放在样板 grSpec 上所需的转动角不大于 α 度，便假设 ES 的点是样板配合的。当发现一点需要大于 α 度的转动时，便将该点分类为不配合的，而结束向左的搜索并在该峰的右侧重复该过程。

图 3.2a 示出带有用于峰之一的样板的能量谱。图 3.2b 中给出将 ES 点放在样板上所需的转动角的余弦。很清楚，对于靠近峰的点的余弦接近 1（小角度）。到接近零的值或甚至负值的改变是相当陡的并与我们直觉地指示为配合区的区的极限很好地对应。在图 3.2c 中将最终铭记为与峰样板很好地配合的点涂上红色。

图 3.3a 示出作为信号 ‘one four zero’ 的时间与位置的函数的耳蜗的能量输出。3.3b 中涂成红色的区为分类为与峰样板配合的部分。

如果确定了与峰配合的一个区，便能为局部 SNR 得出测度。这一测度应给出峰上的能量与配合区结束处的能量之间的比。因为在两端上的值可以非常不同，采用两者的平均值。应指出因为 dB 标度为用于 SNR 的逻辑标度，用于 dB 来确定两边的 SNR 然后求平均的值。图 3.3c 中配合区是用为该配合区确定的 SNR 值颜色编码的。图 3.4 给出噪声条件（汽车工厂噪声，0dB）中这一技术的结果。很清楚，

该方法识别出能获得用于确定在噪声中的语音信号发声部分的频率内容的可靠信息的区。

在本申请中给出脊下的本振频率的确定的描述。因为脊下的信号是近似正弦形状的，并且频率接近本振特征（或最佳）频率 f_c ，可以假设自相关具有余弦形状，其第一最大值在对应于 $1/f_c$ 的延时上，这蕴含可取 $1/f_c$ 作为峰搜索的第一估算值，它能利用例如内插来可观地更快地且比所描述的方法更精确地找到峰。作为可靠性的附加测度，能确定自相关值与假设的余弦形状之间的对应性。

在上一部分中已确定了脊。现在必须确定哪些脊一起属于与/或哪些脊属于目标源。为此采用我们的目标在各时刻（在各 5ms 帧中）上具有唯一地定义的音调这一假设。因此，必须确定哪些脊共用公共的音调。换言之，必须找出能说明大多数找到的脊的音调频率。然后便能在下面的节中利用这一音调估算值作为从背景中选择源的基础。

音调估算利用在前一节中确定的所有脊的频率。这些频率构成的对所谓子谐波求和的输入（见 JASA 83 (1)，1988 年 1 月，“用子谐波求和的音调测定”，D.J.HermeS，中的详细描述）。

子谐波求和运算说明如下。为已找到的各频率，确定这一频率可能已产生的音调值。这些可能的音调值称作子谐波。为了避免大量的可能音调值，尤其是低频范围中，将可能的音调值的范围限制在[45Hz—400Hz]。这一范围跨越在人类话音中发现的音调值。同时将可能的输入频率的范围限制在[70Hz—2000Hz]上。这是能精确地确定的频率的范围。现在有了在当前帧中找到的各个频率的一组可能的音调值。

因为不能以无穷的精度确定频率，在各可能的音调值附近形成高斯分布。这得出在当前帧中找到的各频率（输入频率）的音调候选者的可能性分布。在所有输入频率上求和这些分布。如果音调在大多数分布中具有高概率，它将在求和的分布中得到高概率。如果音调值只在单独分布的一小部分中具有高概率，则会降低其在求和的分布中的

概率。这一过程导致出现在单独分布最大部分中的音调上的高峰。这一点示出在图 3.3 中。

图 3.3A—D 示出单独的频率的音调分布。图 3.3A 中, 采用 175Hz 的输入频率。图 3.3B 中, 采用 350Hz 的输入频率。图 3.3C 中采用 525Hz 的输入频率。图 3.3D 中采用 750Hz 的输入频率。这些频率是 175Hz 的前四个谐波。图 3.3A 中, 最低频率 (175Hz) 在音调范围中只有两个可能的音调: 87.5Hz 与 175Hz。对于图 3.3B—D 中的较高谐波, 可能的音调的数目增加。逻辑上在所有的图 3.3A—D 中可见到 175Hz 上的峰。但是在 87.5Hz 上也在所有四种分布中得到峰。这是因为它正好是 $175\text{Hz}/2$ 。这意味着 175Hz 的所有谐波也都是 87.5Hz 的谐波。它们是在其中所有四个谐波具有峰的范围[45Hz—400Hz]中仅有的两个频率。求和这些分布得出图 3.3E 中的曲线。已经清楚 87.5Hz 与 175Hz 是最佳候选, 但在其它频率上 (诸如在 350Hz 与 116.6667Hz 上) 得到明显的成分。

但是这些音调候选者只说明输入频率的一部分。因此, 将各音调候选者乘以它所说明的输入频率的分数。这得出图 3.3F 中的曲线。显而易见分布中的额外的峰明显地减少了。然而, 仍留下两个候选者: 87.5Hz 与 175Hz。因为选择了 175Hz 的谐波, 希望它是最佳的得出的候选者。达到这一目的的方法是通过“惩罚” 87.5Hz 上的候选者, 因为它会在 262.5Hz、437.5Hz、612.5Hz 等上产生中间谐波但这些频率不在输入中这一事实。“惩罚” 87.5Hz 上的候选者因为其谐波的一半从输入中丢失这一事实。因为这里对 175Hz 与 87.5Hz 发现的情况频繁地出现并导致所谓八度误差。对丢失谐波的惩罚更严厉 (二次的, 这可视为这一校正的双重作用)。图 3.3G 中的曲线中所示的最终分布现在在 175Hz 上显示高峰并且没有带有高概率的实际代替品, 这正是我们所希望的。

当我们采用下述频率时情况稍为复杂: 250Hz、375Hz、500Hz

及 750Hz。得出的曲线以类似于图 3.3A—G 中的曲线的次序示出在图 3.4A—G 中。这些频率可能是 125Hz 的二次、三次、四次与六次谐波。在这一情况中，我们会有相当大的数目的丢失的谐波。或者我们会有来自某个其它源的 250Hz 的一次、二次与三次谐波及偶然的 375Hz 分量。在图 3.4E 中，能见到子谐波求和的结果。很清楚在本例中存在音调的两种选择：125 Hz 与 250 Hz，并且这两种选择是大致均等地正确的。

在这种情况下，必须利用来自其它帧的附加信息才能在这两种选择对象中作出选择。如果在后面或前面的帧中发现 125 Hz 的其它谐波，但选择这一选择对象。然而，如果只找到 250 Hz 的谐波并且 375 Hz 分量具有与其它分量不同的时间发展，则可以得出它是由不同的源产生的结论。

因为确定当前帧中所有频率分量的子谐波的分布的计算是繁杂的，我们选择构成用大量（550）输入频率的子谐波填充的矩阵。这称作音调矩阵并能在初始化时计算。通过在具有最接近输入频率的频率的两行之间的线性内插能够容易地从这一矩阵确定输入频率的分布。图 3.5 中示出该矩阵及组合 175、350、525 与 700 Hz 的实例选择的行，连同得出的最终分布。

无噪声信号的基本周期估算

已研制了可应用于无噪声（语音）信号的基本周期估算设备。本节中提出的算法旨在作为前面提出的更一般的周期估算技术的快速替代品。这一设备可用于在自动语音识别之前为语音数据库估算基本周期轮廓。

对测定语音识别系统的坚韧性的基本周期估算算法的需求稍为不同于旨在选择与跟踪尽可能多的源的系统。这是一般基本周期估算的优化。

对于 ASR 测试，有必要生成尽可能好地与存储的样板相似的信号表示，这意味着噪声应尽可能少地污染这一选择。起始期间，但在停止期间更普遍，信号能量可能是相对地低的，同时不传递语言学信息。例如， $t=360\text{ms}$ 之后图 2.2 中的字/NUL/中的信息没有意义，而能估算上升的音调至少另外 100ms。在这最后的 100ms 期间，信噪比急剧下降而得出更污染的 TAC 选择。为了减少污染在确定信号的起始或停止时是否发音时保守一点是有利的。

这是通过限制耳底膜模型一半的低频中的脊的能量与衰减表现两者来实现的。耳底膜的这一部分很少受到未发声的信号分量的影响。当在 10ms 中能量损失对应于 50%或以上时，或者当能量不超过发音的预期最大能量的 1%时，便认为该帧是未发音的。这一组合的标准可用更完善的标准替代。

衰减标准比在没有输入时的泄漏积分过程的衰减更具限制性。与 10ms 的泄漏积分时间常数关联的 10ms 中的衰减为 $e^{-10/t}=e^{-1}=0.37$ ，而应用的阈值为 0.5。对于语音信号这一阈值是非常高效的。因为基本周期轮廓估算是好的并且因为在信号中没有驱动能量，基于继续的基本周期轮廓的选择通常几乎与基于连续型式的周期轮廓的选择没有区别。两种标准的组合导致趋向于具有较早的停止的基本周期轮廓，正如 ASR 应用所要求的。

基本周期算法是基于沿脊的自相关的求和的。这一算法类似于声称模拟人类音调感觉特征的基于相关图的算法。主要差别是脊下的运行自相关的使用而不是计算及求和基于 FFT 的相关图。这减少了计算负担并因为不应用准平稳性而对于急剧地改变音调的信号能达到瞬时音调的更好的估算。如在图 3.7 的上下文中所提到的沿来源于同一源的脊的自相关与作为所有的脊共用的第一公共周期性的基本周期一致。图 3.7 示出一组自相关及对应的求和的典型实例。注意所有自相关是简单地相加并且未进行群时延校正。因而结果是近似的。选

择算法(见周期性信号成分的选择)中的优化确定最终瞬时基本周期。

在各帧中选择具有高于 0.3 倍的沿脊的局部能量的值的求和的自相关中的三个最高峰,并按照自相关值从最高的开始排序。当没有峰满足这一标准时,认为该帧是未发声的。假设这些自相关滞后之一对应于这一帧所要求的基本周期值。将所选择的峰组合到时间轮廓中。丢弃持续时间小于 25ms 的轮廓。在各帧中,将剩下的轮廓与最低的脊的段的对应最佳频率比较。选择在 60%或以上时间中落入最低的脊的最佳频率的 10%以内的轮廓,丢弃其它轮廓。最后,用噪声中的基本周期估算节中所描述的相同过程平滑所选择的周期轮廓。算法的最终输出为符合周期性信号成分的选择节中的 TAC 选择算法的要求的定义各帧上的本振瞬时基本周期的一阶逼近的参数。

这一技术组合互相补充的两种知识源:自相关中的周期性信息提供精确的周期性估算,及位置信息方便了正确的基本周期候选者的选择。

周期性信号成分的选择

下一步骤是将信息实际赋予听觉事件。图 2.17 下方示出基于 TAC 的听觉事件的典型实例。单一制约的应用,周期轮廓在上方图中的有噪声的信号上已具有的有利的结果。在低频侧,TAC 耳蜗线图可靠地选择基波,在高频侧,它将大面积的时间-频率(实际上是时间-位置)平面分配给听觉事件。在低频侧,所选择的区是受单一谐波控制的。在高频侧,这些区是受共振峰控制的:符合公共基本周期的谐波复合体。

因为小的音调估算误差能导致大的后果,TAC 选择最好是基于局部优化过程的。音调估算算法所估算的基本周期轮廓很好地描述周期轮廓的总的发展,但并不代表快速音调波动。估算过程为各帧得出周期值及局部时间导数。这给出局部时间发展的可靠指示,但实际的瞬

时周期可在这一平均发展附近波动。本振周期估算及其导数得出本振瞬时周期性值 T_s 。图 2.18 中的虚线作为本振瞬时周期性曲线示出这些值。为了估算瞬时周期的最佳值，在周期性上向上及向下位移这一曲线（即在图 2.18 中右与左）并计算本振瞬时周期的各种选择的对应 TNC 值。选择使得压缩的 TNC 的正值之和为最大的瞬时周期的选择作为以之为根据的最终瞬时周期的选择。

最佳瞬时周期是最优化正值与 X 轴下面的区的值。这一最优化过程是降低自然音调波动的效应与小周期估算误差的非常高效方法。

虽然图 2.17 中描绘的选择是基于正确的周期轮廓的，它不能保证选择是正确的：背景说话人之一可能是周期轮廓之一的源。利用诸如说话人的话音类型及语言的所有方面等可能的源类型的知识的进一步处理必须解决这一问题。幸运的是，基于在相当嘈杂的情况中估算的正确的周期轮廓的听觉事件所表示的信息包括关于单独的谐波与共振峰的相对重要性的精确信息。这足以将数据的可能的解释的数目减少到少数假设。

虽然 TAC 方法不能将非周期性信息赋予听觉事件，它能帮助确定有可能分配给同一流的非周期性听觉事件的可能候选者的位置。在正常语音中非周期性信号分量的位置强烈地与周期性分量相关。在大多数情况中，这些成分正好在起始之前或起始期间结束，并在周期性成分的停止期间或之后开始。在/TWEE/（/TWO/）的/T/的情况中，开始在 $t=1000\text{ms}$ 时并在图 2.17 的上方图中从 90 到 100 的段范围中最明显，样板匹配的简单方式可足以检测非周期性成分的可能候选者。

特征周期相关（CPC）

为了对付象起始与非周期性噪声等非周期性信号，发明了其它设备。虽然周期性信号是以信号成分的离散集合为特征的，非周期性信

号是以频率的连续分布为特征的。诸如单位脉冲等典型的非周期性信号表示频率的相等地加权的分布。频率的每一个范围试图用对应的特征频率 f_c 引开耳底膜段的范围。这导致在其中 BM 的各段以其特征频率的平均数振荡的情况。这导出特征周期相关 (CPC) 的初始定义如下:

$$r_s^c(t) = L(X_s(t), X_s(t+T_{c,s})) \quad (27)$$

其中 $T_{c,s}$ 表示段特征周期。CPC 为用特征周期振荡的耳底区提供接近耳蜗线图的能量值的值。它还为被以比特征周期大得多的周期振荡的低频分量引开的区产生高值。为了保证 CPC 对后面的区较不敏感, 可将 CPC 重新定义成:

$$r_s^c(t) = \frac{L(x_s(t), x_s(t+T_c)) - L(x_s(t), x_s(t+T_c/2))}{2} \quad (28)$$

它对应于 $X_s(t)$ 与 $X_s(t+T_{c,s})$ 之间的相关与 $X_s(t)$ 与 $X_s(t+T_{c,s}/2)$ 之间的相关的差的一半。当段以接近特征周期的周期振荡时, 后一相关将是负的, 但其绝对值是与第一相关相当的。这两个相关之间的差的一半使得 CPC 值接近耳蜗线图值。对于以大于 2 倍特征周期的周期振荡的段, 得出 $X_s(t)$ 与 $X_s(t+T_{c,s}/2)$ 之间的正相关, 这与耳蜗线图的值相比将 CPC 降低到小的值。有可能通过组合依赖于 T_c 的不同分数的成分公式化显示类似特性的 CPC 的其它实施例。

可在设备中利用 CPC 来估算非周期性 (及周期性) 信号分量。根据其定义, CPC 极大地适应以接近特征频率的周期振荡的段。可用诸如下述标准识别这一点在其中成立的区:

$$r_s^c(t) > C_s r_s(t), \quad C_s < 7 \quad (29)$$

其中 C_s 为取决于段号的常数。对于所有的 s 值, 这一常数的典型值是 0.85。满足这一标准的段-时间平面的区是被带有对应于特征频率的频率的信号分量引开的。注意 CPC 提供确定非周期性信号分量的有用方法, 但也能有效地用于提高识别被周期性信号分量补充的

区的精度。可将它用在可靠性测度的估算中。

起始与停止检测

CPC 以及耳蜗线图能用于精确定位起始与停止。它们是经定义为非周期性的，并得出频率分量的闭联集。起始可与上升时间不同。正弦曲线的快速起始导致带有在少数局部群时延 d_s 内发展的非常宽的频率内容过渡到理想的正弦响应。正弦曲线更平缓的建立导致与理想的正弦响应的强度逐渐建立结合的减小的过渡。当正弦曲线的上升时间比局部群时延小时，过渡表示得出明显的脉冲样响应的宽广频率范围。当上升时间比局部群时延大时，过渡是小的或不存在的。在起始后的各时间点上，理想的正弦响应近似于 BM 响应。注意停止出现在与老信号破坏性干扰的新信号起始时。因此，检测停止的技术非常类似于检测起始。因此将焦点集中在起始上。

检测起始的方法之一是通过测定来源于耳蜗线图或 CPC 之一的逐渐变高的能量梯度是否超过阈值：

$$\frac{\frac{\partial}{\partial t} E(s, t)}{E(s, t)} > C_{onset}(s, t_0) \quad (30)$$

能使用的阈值 $Conset(S, t_0)$ 的实例为 n 倍 $[t-t_0, t]$ ms 时间间隔中的能量变化（对于有噪声的语音 $n=2$ 及 $t_0=20$ 工作得很好）。可用公式表示依赖于局部群时延与/或依赖于段方向 S 上的梯度的其它阈值。

在大多数应用中，识别起始的信号分量的类型是重要的。这能用更仔细地分析信号分量的起始与连接的发展的细节来达到。初始过渡的宽度（用响应的段的频率范围表示的）与持续时间结合稳态信号的上升时间是信号的上升时间的可靠指示。在语音情况中爆破音/B/、/P/与特别是/T/产生包含大量的段的过渡。这些在起始的图形描绘中显露垂直的结构。在诸如/S/与/f/这样的噪声脉冲串中这些宽过渡是丢失

或最少的。发出的语音的起始与局部群时延比较通常是慢的并且起始效应是最小的。在语音的非特征性快速起始（与停止）的基础上，能容易地识别诸如电话铃声等人工声音。

屏蔽构成

对于象自动语音识别与语音编码等应用，识别符合要识别的信号类型的特征要求的信号分量是重要的。耳蜗线图、CPC、及 TAC 选择与对应的周期轮廓的组合可用于识别屏蔽：即反映符合目标类型的要求的信号分量的段-时间平面中的一组区域。这一屏蔽定义在语音编码或语音识别期间将处理哪些信号分量，因此是具有关键重要性的。

构成屏蔽是两阶段过程。在阶段 1 中单个地描述所有信号分量并丢弃不符合信号的目标类别的特征要求的信号分量。在阶段 2 中互相结合评估所有接受的信号分量并作出最终选择。因为屏蔽构成关键性地取决于目标类别的特征特性，不可能公式化完整的标准组。本节的其余部分讨论用语音信号处理作为实例的一些有用的标准。

阶段 1 的信号分量评估包含单个信号分量的识别。这一阶段是通过排除具有不足的能量的小区或能量在接近与低通滤波关联的速率的速率上下下降的小区来进行的。当用泄漏积分器实现低通滤波时，它在不存在 τ ms 到 $e^{-1}=0.37$ 原始值内的输入时下降。供语音排除以对应于每 τ ms 损失一半能量的速率衰减的所有信号分量的适当的阈值。这些区域的排除可实现为：

$$m = m_{org} \left\{ E(s, t) > C_{Energy}(s, t), \frac{\frac{\partial}{\partial t} E(s, t)}{E(s, t)} > C_{decay}(\tau) \right\} \quad (31)$$

Morg 为整个段-时间平面， $C_{Energy}(S, t)$ 为能量的阈值。这一阈值可以是绝对值或取决于局部能量的长期平均值的值。 $C_{decay}(t)$

为衰减速率的阈值。

在准周期性信号的情况中，能估算基本周期轮廓及计算 TAC 选择。为了识别 TAC 选择内的最重要的谐波成分及丢弃由偶然的相关引起的大多数虚假成分，可将 TAC 选择与耳蜗线图的能量值比较。每当 TAC 选择超过局部能量的一定分数（诸如 0.25）时，便接受对应的时间-段区。这得出用下式定义的屏蔽 M_{TAC} ：

$$m_{TAC} = m\{r_{S,T}(t) > C_{TAC}(S, t) r_S(t)\} \quad (32)$$

M 表示整个段-时间平面， $C_{TAC}(S, t)$ 为与 S 及 t 相关的接受阈值， $r_{S,T}(t)$ 及 $r_S(t)$ 分别表示 TAC 选择及耳蜗线图。注意各周期轮廓 $T(t)$ 得出一 TAC 屏蔽。

受匹配的频率分量引开的区可用下式标识：

$$m_{CPC} = m\{r_S^C(t) > C_{CPC}(S) r_S(t)\} \quad (33)$$

其中 $r_{CS}(t)$ 表示例如计算的 CPC， $C_{CPC}(S)$ 为局部引开的依赖于段的标准。

起始可用类似的策略标识：

$$M_{Onset} = M\left\{\frac{\frac{\partial}{\partial t} E(s, t)}{E(s, t)} > C_{onset}(s, t_0)\right\} \quad (34)$$

其中标定的能量梯度（用 CPC 或耳蜗线图计算的）必须超过取决于沿 BM 的位置及依赖于 t_0 的某一时间范围的阈值 $C_{onSet}(S, t_0)$ 。

注意起始屏蔽通常是 CPC 屏蔽与/或 TAC 屏蔽的子集。TAC 选择的峰同样显露在 CPC 屏蔽中。在组合屏蔽“下面”表示的信息提供能用于从进一步处理中排除屏蔽区的大量基本信号特性。这需要关于目标信号的知识。从进一步处理中排除一些区的典型的语音标准包括：

- 一对应于超出正常人类音调范围的音调的 TAC 选择
- 一帶有（太）恒定的音调的 TAC 选择

一帶有不现实的共振峰模式的 TAC 选择

一帶有不现实地快速的起始的 TAC 选择

缩减（例如对于高频分量 20ms 及对于低频分量 40ms）或进行的 TAC 屏蔽的相关区域包含少于两或三段。

通过应用两种互补的方法非周期性成分的估算是可能的。第一方法按照下式从 CPC 屏蔽排除用 TAC 屏蔽表示（每一周期轮廓 $T(t)$ 一个）的区：

$$m_{\text{Aperiodic}} = m_{\text{CPC}} - m_{\text{TAC}} \quad (35)$$

这得出非周期性信息的屏蔽。这一屏蔽能用类别特定的标准裁剪成只显示段-时间平面的足够面积的相关成分。

另一方法是为 CPC 屏蔽的主相关区计算运行自相关。当该运行自相关在少数局部特征周期内下降到接近零的值时，该区是非周期性的。这一方法可用于断定在第一遍中估算的区是真正非周期性的。这两种方法的结合得出反映非周期性区的段-时间平面的相关区的可靠估算。

在屏蔽构成的第二阶段中，组合单个的信号分量。对于这一阶段，必须使用其它依赖于目标类别的限制。对语音的典型限制是：

一语音信号的不发音的分量是强烈地与语音的发音部分相关的。可以丢弃孤立的发音的信号分量。

一在平均音调上相差太大的发音分量不太可能来源于同一源。接受的发音信号成分应构成语言学上似乎合理的周期轮廓。

一重叠的音调轮廓不能来源于同一源。

这些限制的应用得出一或多个屏蔽，它们表示可能来源于目标源的不同信息形式。最好地利用这一信息便是进一步处理阶段的任务。

注意用于建立屏蔽的所有变量具有关联的可靠性测度。在屏蔽构成过程中可利用这些可靠性测度。

逆耳底膜滤波

因为 TAC 为将信息赋予听觉事件构成可靠的基础，有人可能感到奇怪这是否能用于将组合声音分裂成组成的声源。

控制 TAC 耳蜗线图的时间-位置平面的一定区的所有准周期性信号成分表示耳底膜振动。由于耳底膜模型的这一实施例是作为基于脉冲响应的有限脉冲响应（FIR）滤波器实现的，有可能通过在时间上倒置脉冲响应及补偿由耳底膜滤波器的双重使用所导致的频率效应来逆向滤波。

全逆向得出原始的信号混合物。但如果逆滤波是基于被目标源补充的时间-位置平面的区的，则输出是理想地只基于目标的信息的。上节中所定义的屏蔽可用于这一目的。为了减小起始与停止的效应，将屏蔽裁剪成包含单个段的长的连续成分：填满 TAC 轨迹的正值中的小洞及丢弃孤立的正点。最后向屏蔽提供平滑的 10ms 宽的起始与停止。这得出图 2.19 下方图中描绘的屏蔽。

为了改进声音质量，不完全丢弃背景，而是用一可调节的因子减小：在本例中振幅为 100 的因子（用能量表示为 40dB）。通过不完全丢弃背景，减少了不自然的“深度”寂静及非周期性成分的某些迹象，诸如/TWEE/的/T/，保留在信号中，这方便了感觉。当将得出的重新合成的声音再一次提交给耳底膜模型时，便能计算重新合成的声音的耳蜗线图。这是在图 2.20 的中间的图中提供的。这一重新合成不包含不发音的区，但可将这些区加到屏蔽上。然而可靠地组合发音与不发音的区非常困难。

上方图示出原始信号的耳蜗线图。这一信号构成信息的唯一的源：没有用也不需要先验的信息。下方图构成洁净的对照。除外最后一个字的第二共振峰结构是完全受到屏蔽之外，忠实地表示了所有重要的周期性成分。注意重新合成的耳蜗线图更“模糊”，这是由背景的虚假成分引起的。避免这一点的方法是测定与平滑所有单个信号分

量并在真实的语音合成过程中将它们加在一起。

重构耳蜗线图

前一节中计算的重新合成的声音可用作诸如象 PhilipS FreeSpeech 与 L&H Voice XPress 等语音识别软件等标准识别系统的输入。识别系统的替代的及计算上有利的输入可以是基于 TAC 耳蜗线图的。

对于 ASR 系统的适当的输入为描述目标语音的频谱包络的时间发展同时抑制音调的效果的图。如图 2.7 的右上方图所证明的洁净信号的发音部分的 TAC 耳蜗线图紧密地与标准耳蜗线图相似。通过填满负值可将图 2.17 的下方图中的 TAC 耳蜗线图裁剪成更好地与洁净耳蜗线图相像。

这一重构过程为图 2.21 中所示的两步过程。第一阶段搜索单个谐波的证据并利用这一证据来计算重构的下一半。第二阶段利用屏蔽及不同屏蔽效应的逼近加上关于频谱的其余部分的信息。

算法的第一阶段是在选择的前 60 段中估算相关的脊。如在基本周期估算算法中, 这些脊是通过组合相差小于 2 个段号的接连的峰构成的。接受长于 15ms 的脊作为谐波的候选者。由于基本周期轮廓是已知的, 有可能预测基波的段号。接受平均在前 4 个谐波的预期值 1 段以内的脊作为谐波。注意这一标准能在时间发展上失配的基础上丢弃虚假的脊。能以这一方法建模的谐波数取决于耳底膜的空间分辨率。具有较锐利地调谐的 BM 模型及较高数目的段, 便能单个地建模较高数目的谐波。在本例中单独地处理 4 个谐波, 因为前 4 个谐波的接受区并不在当前的 BM 模型中重叠。本算法对这一参数的值只是弱敏感的。

图 2.21 的左上图示出所有候选的脊。通过用三点局部平均值取代各值来平滑能量沿这些脊的发展。通过加上符合式 6 的接连的谐波

的成分利用这些平滑后的谐波脊来重构原始耳蜗线图的估算。这一过程示出在图 2.22 的上方。

重构从加权基频的理想正弦响应（如图 2.5 中所示）开始。假设谐波只在频率中向上互相影响。在二次谐波的位置上，一部分能量起因于基波，而其余能量用于加权二次谐波的理想正弦响应。图 2.22 中，三次谐波的位置的大部分能量必然是起因于二次谐波的，四次谐波相对地更重要。只利用 4 个谐波得出的部分重构用黑色描绘。

该算法的第二阶段是重构高频范围。再一次利用屏蔽来精确定位最有可能表示目标的信息的区。用在屏蔽下选择的超过部分重构的值取代该部分重构的值。这一步骤的结果描绘在图 2.21 的左下方图中。这一阶段得出带有不现实的向上与向下斜率的高频成分。图 2.22 的上方图中的黑色峰清楚地示出这一点。为了不增加额外的信息而使重构更现实，可用表示与激励脉冲波前（flank）后面的峰的位置的源一致的屏蔽效应的脉冲波前扩大屏蔽的脊。这些再一次能从正弦响应估算并加在重构上。最后，BM 的呼出效应及泄漏积分的效应可建模成指数衰减。这模拟重构的向前屏蔽效应。最终的重构示出在图 2.21 的右下方图中并在图 2.22 的下方图中用黑色画出。

视觉检查显示重构通常具有高质量。诸如图 2.22 的高频范围等一部分信号具有非常不利的局部信噪比。如能在图 2.22 的下方图中见到的，当短划线目标接近对应于总能量的点划线时正确重构的概率是高的。这对应于局部 SNR 有利（ $\text{SNR} > 3\text{dB}$ ）的情况。当短划线与点划线之间的距离增加时，重构不正确的概率增加。当距离大于 3dB 时，这对应于 0dB 的局部 SNR，重构很可能包含虚假成分。可参见段 71 附近的实例。

这得出图 2.23 中描绘的合成相关图。上方图示出基于洁净信号的 TAC 的“重构”。与图 2.20 的下方图的比较揭示这两个图的主要分量非常相似。这表示重构方法的有效性。图 2.23 的下方图示出基

于从有噪声的数据估算的 TAC 的重构。由于这一信号的一部分是屏蔽的并增加了背景的一些虚假成分，匹配是不完善的，但两个图的主要特征是相似的（在视觉检查下）。

这里描述的过程不是达到来自单个源的信号的重构的唯一可能方法。如在各式各样的子过程的描述中所提到的，替代方法是可以得到的。这些替代方法可包含使用可靠性节中所描述的可靠性测度来指导屏蔽的构成，或用逆矩阵法来确定单个谐波的强度。

重构过程的另一实施例采用估算单个信号分量节中所描述的信号分量估算技术的变型。在这一实施例中利用屏蔽来精确定位可靠的区。采用了 E 与 R 两者从而将屏蔽外面的值设定为零。 $W=R^{-1}E$ 的解给出要求的权重。

这一重构过程也能用来描述选择的信号的谐波内容。这一描述可用来例如高效地编码表示该信号的数据。这一谐波内容描述也能用在进一步信号分析中。

参数化

基于 HMM 的 ASR 系统要求不带诸如音调信息等分散注意力的事物的目标语音的频谱包络的估算。图 2.23 中所描绘的图不是非常适用的因为最有活力的分量为基波。虽然它们携带共振峰信息，第一共振峰的详细实现强烈地依赖于音调。为了降低无关的音调差的影响及强调第二与第三共振峰，可将压缩的耳蜗线图的值乘以与段相关的因子。对于第一段这一因子为例如 1，而对于最后的段为 5。中间的段的乘法因子可以是这两个极端值之间的线性内插。

这是与预加重具有类似效果的运算，预加重是在 ASR 标准方法学内通常应用的一种高通滤波形式，并得出所有频率都贡献类似的能量的频谱。这一形式的预加重可跟随估算频谱包络的算法，例如通过连接接连的谐波的峰。

作为最后的步骤，必须尽可能高效地编码耳蜗线图的包络。为了生成一组类似于 MFCC 值的参数，可执行“增强”的耳蜗线图的余弦变换。结果是倒频谱的变量。保留表示低空间频率的倒频谱图的前面的 8 到 14 个值而丢弃其余的值。最终，通过平均接连的值将接连的帧之间的时间步长从 5ms 增加到 10ms。这将帧步长带到标准值并加速处理。将这些值存储在硬盘上并用作语音识别系统的输入。

存储的参数并不是饱含信息量的，但利用逆余弦变换能将它们变换回耳蜗线图样表示。结果示出在图 2.24 中。

两个图都反映语音识别系统能获得的信息。上方图是基于原始洁净信号的。每一段的能量分布是用 1 与 5 之间的值增强的，频谱包络是用 12 个倒频谱系数编码的。与图 2.24 的下方图相比，高频段更明显，基波较不明显，而共振峰特征较宽。下方图是基于图 2.23 的重构的 TAC 耳蜗线图的，它与理想的耳蜗线图具有好的总体符合，但由于屏蔽与虚假的背景成分而是有噪声的。这两种表示适合于作为识别系统的输入。

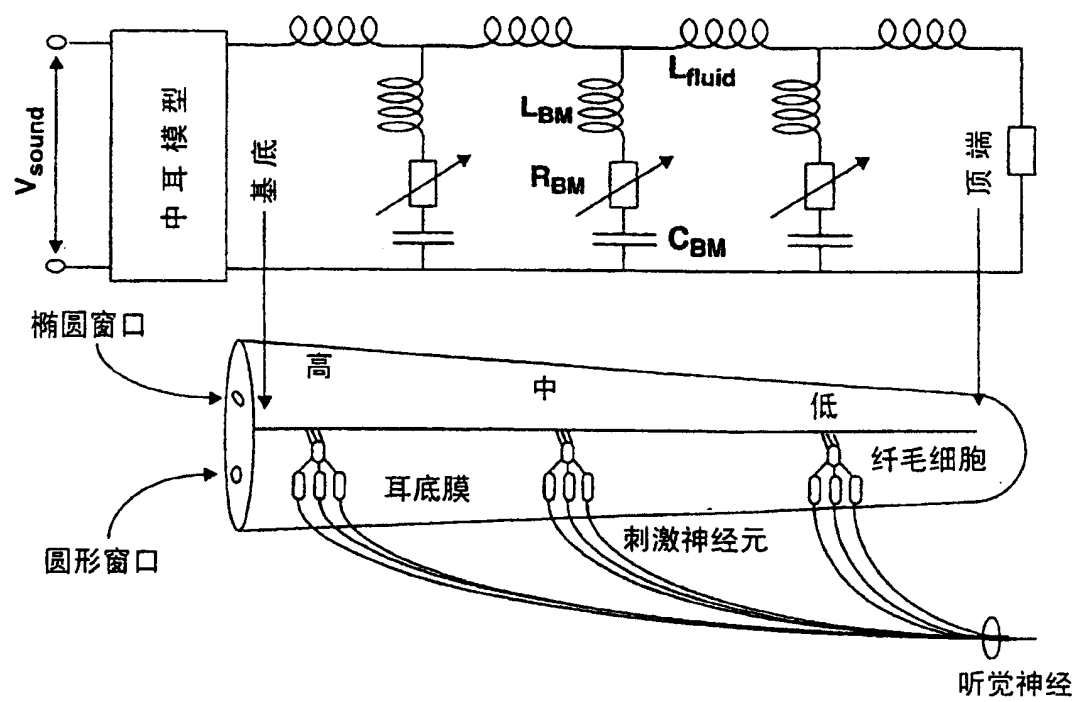


图 1.1

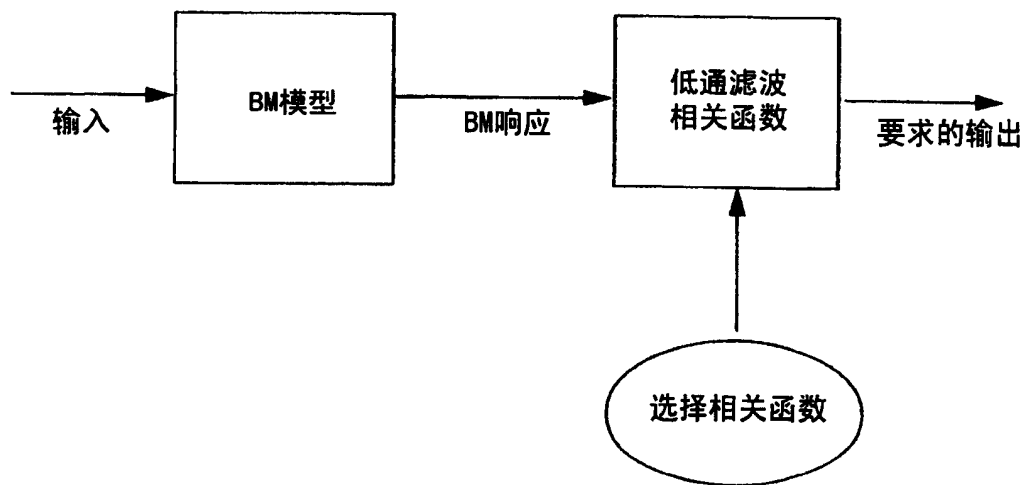


图1.2

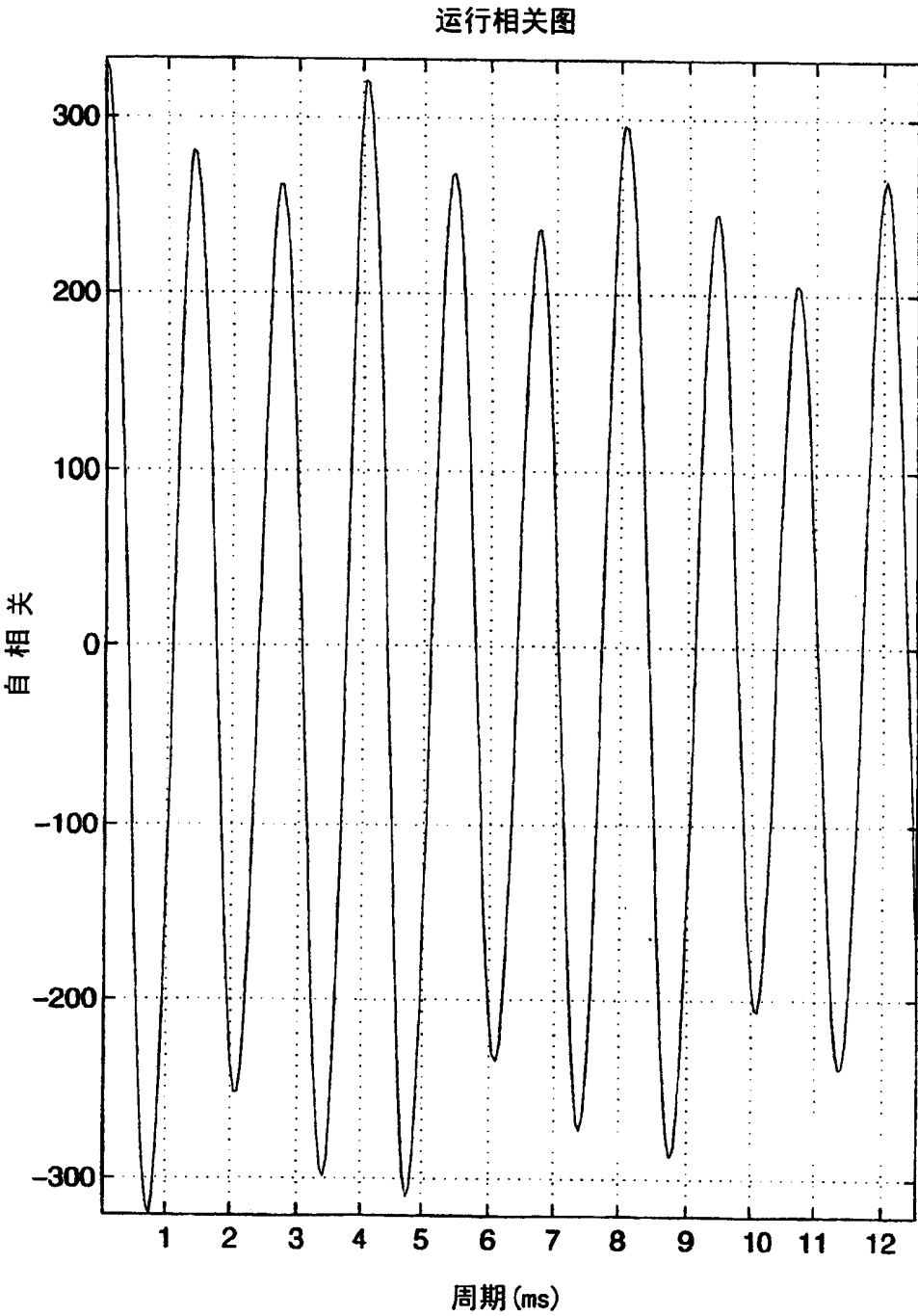


图1.3

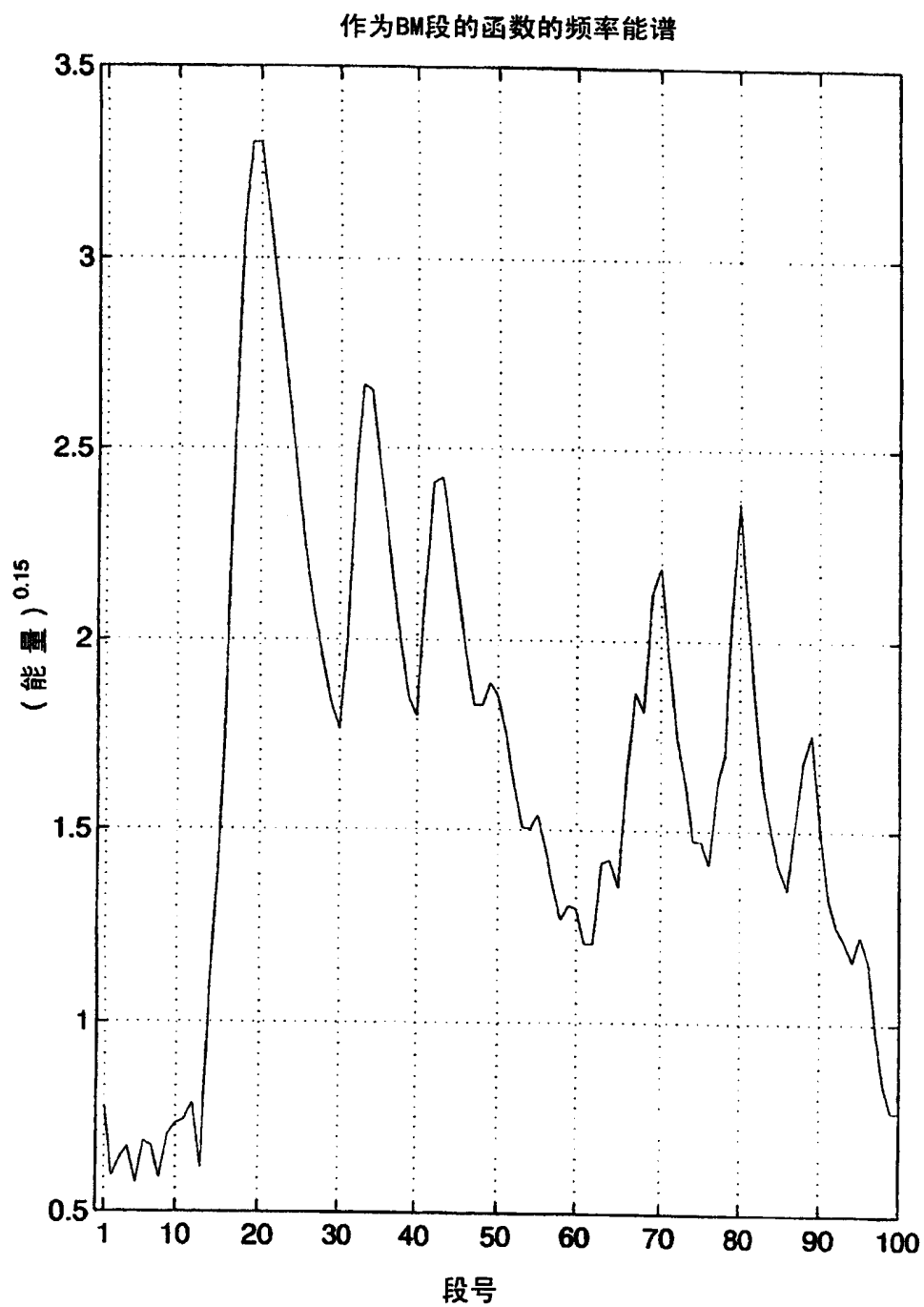


图1.4

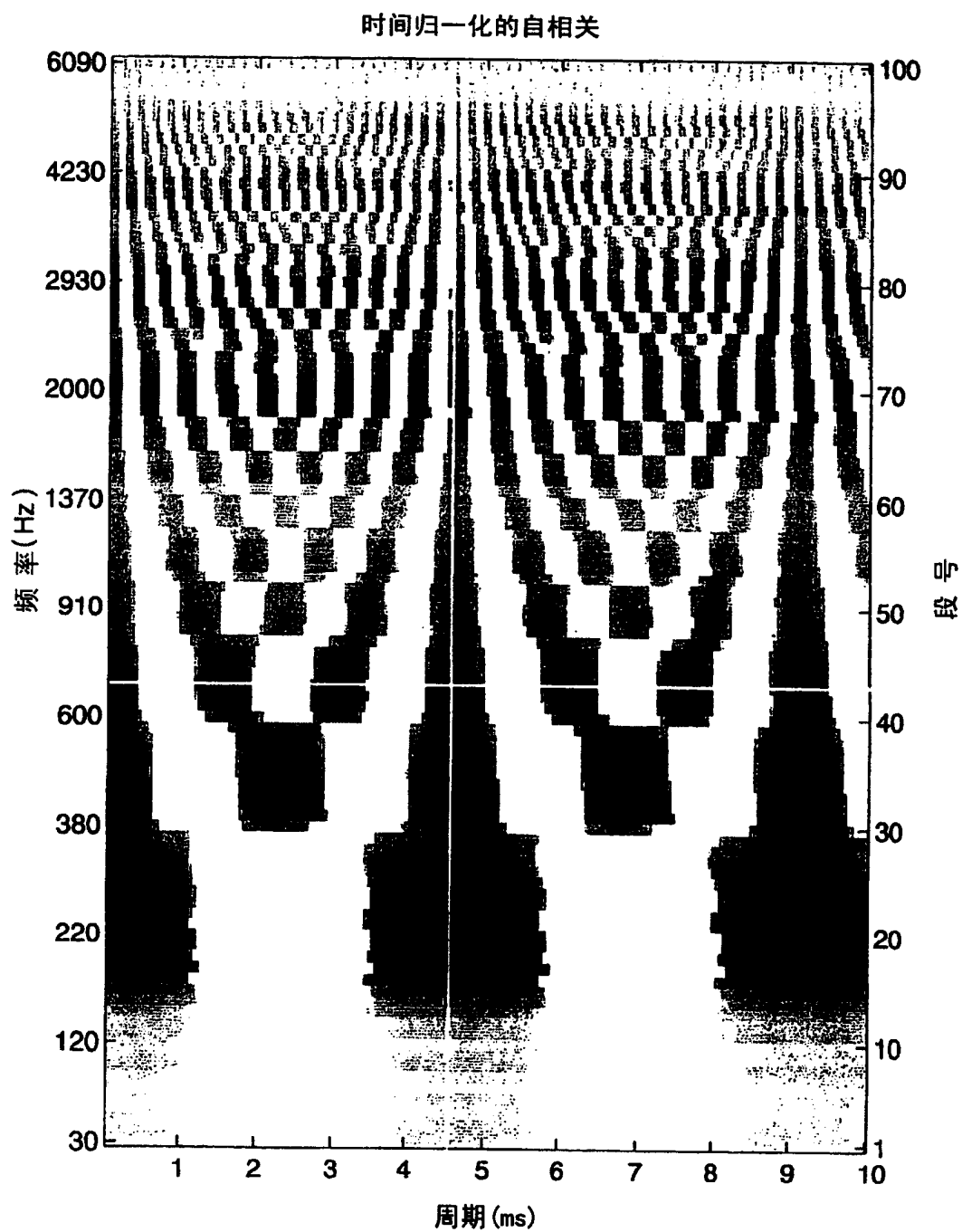


图1.5

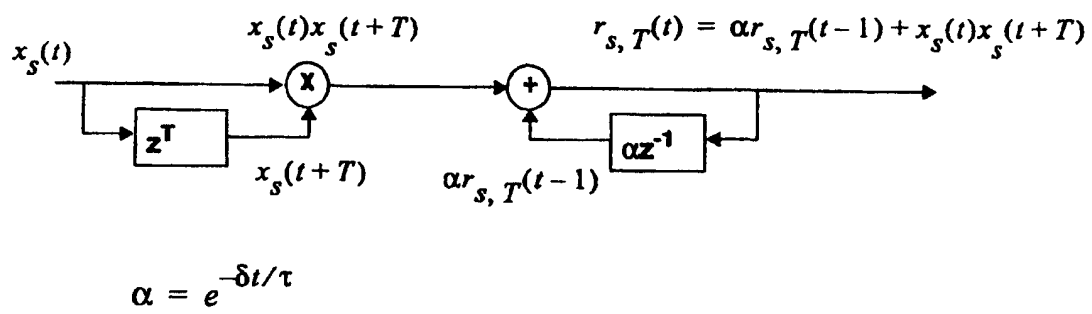


图1.6

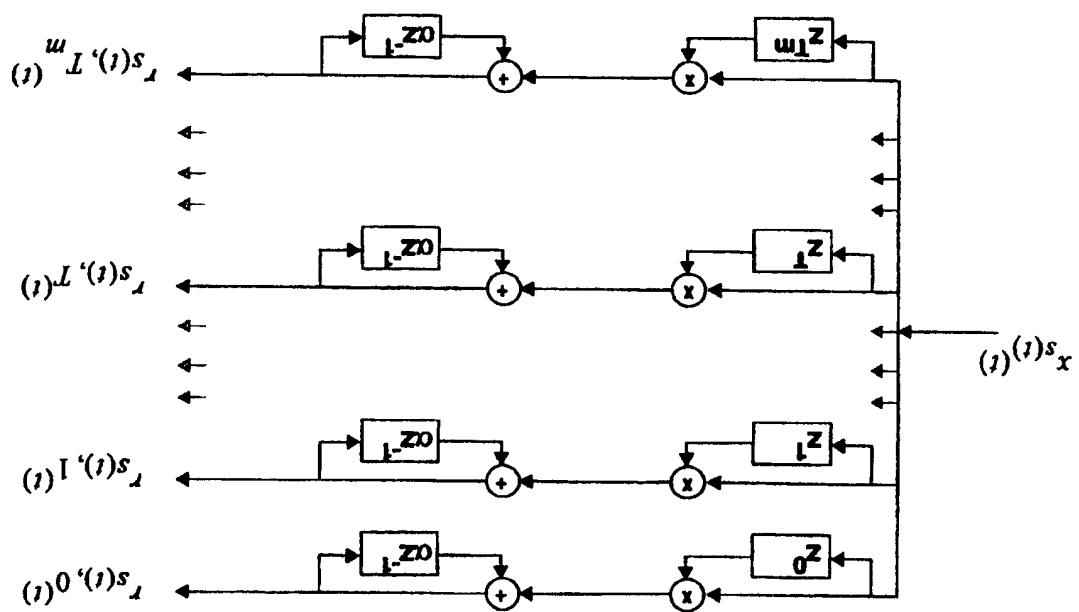


图1.7

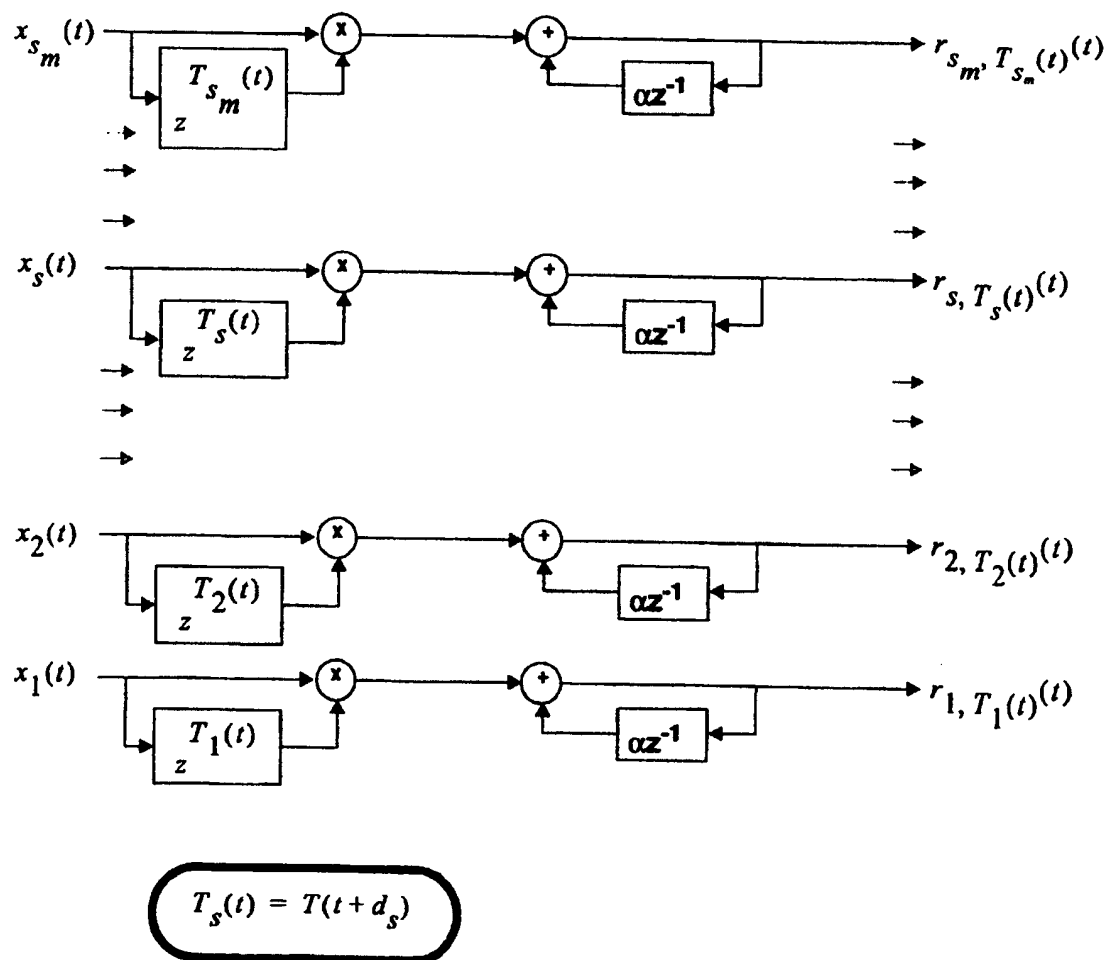
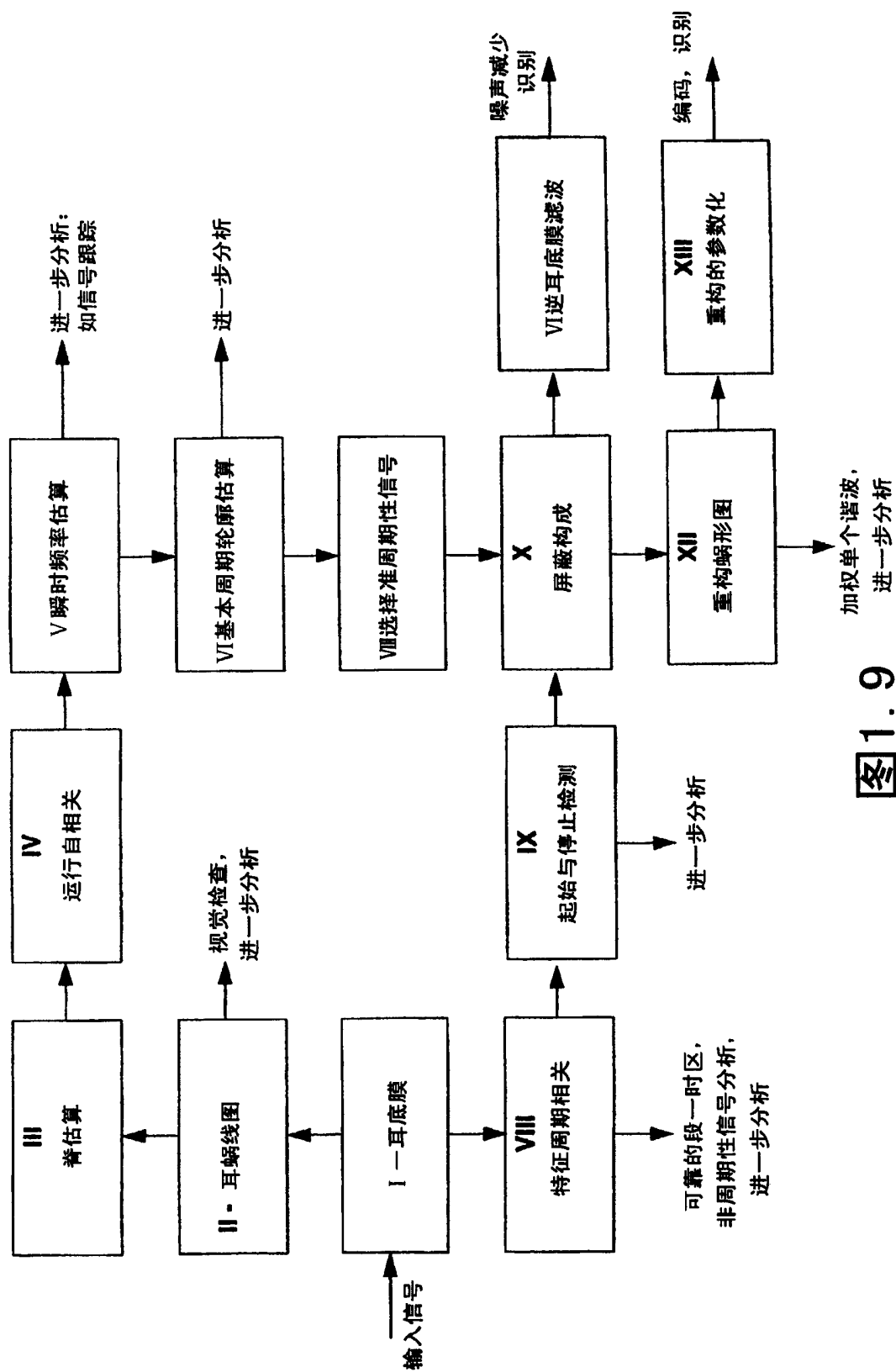


图1.8



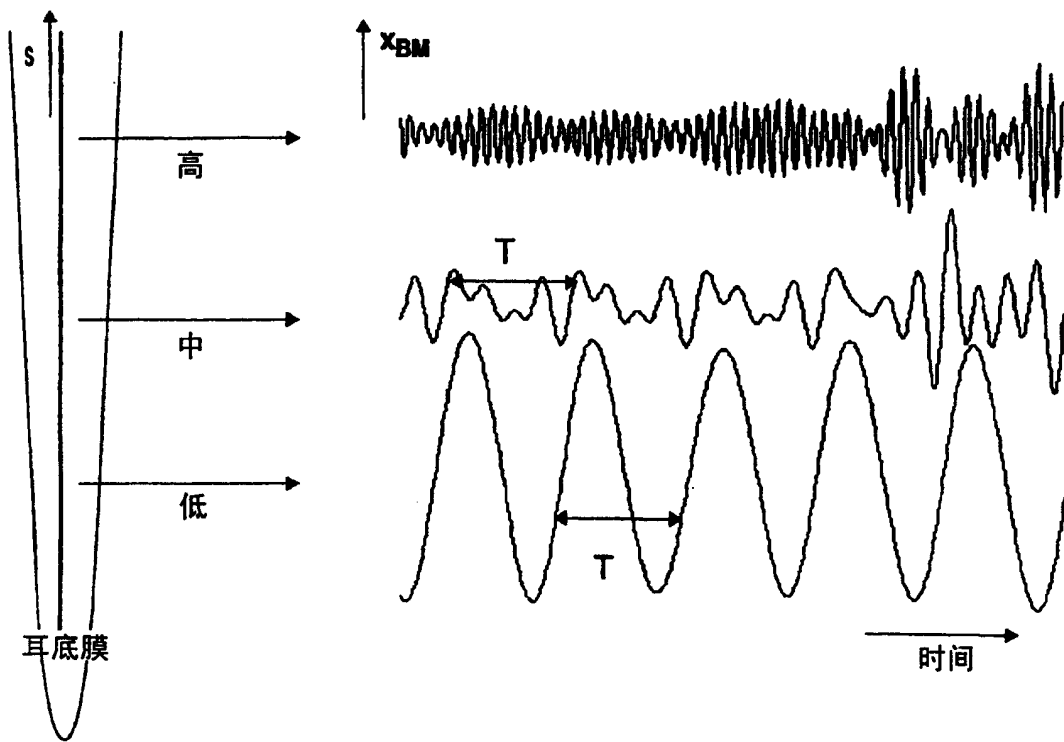


图2.1

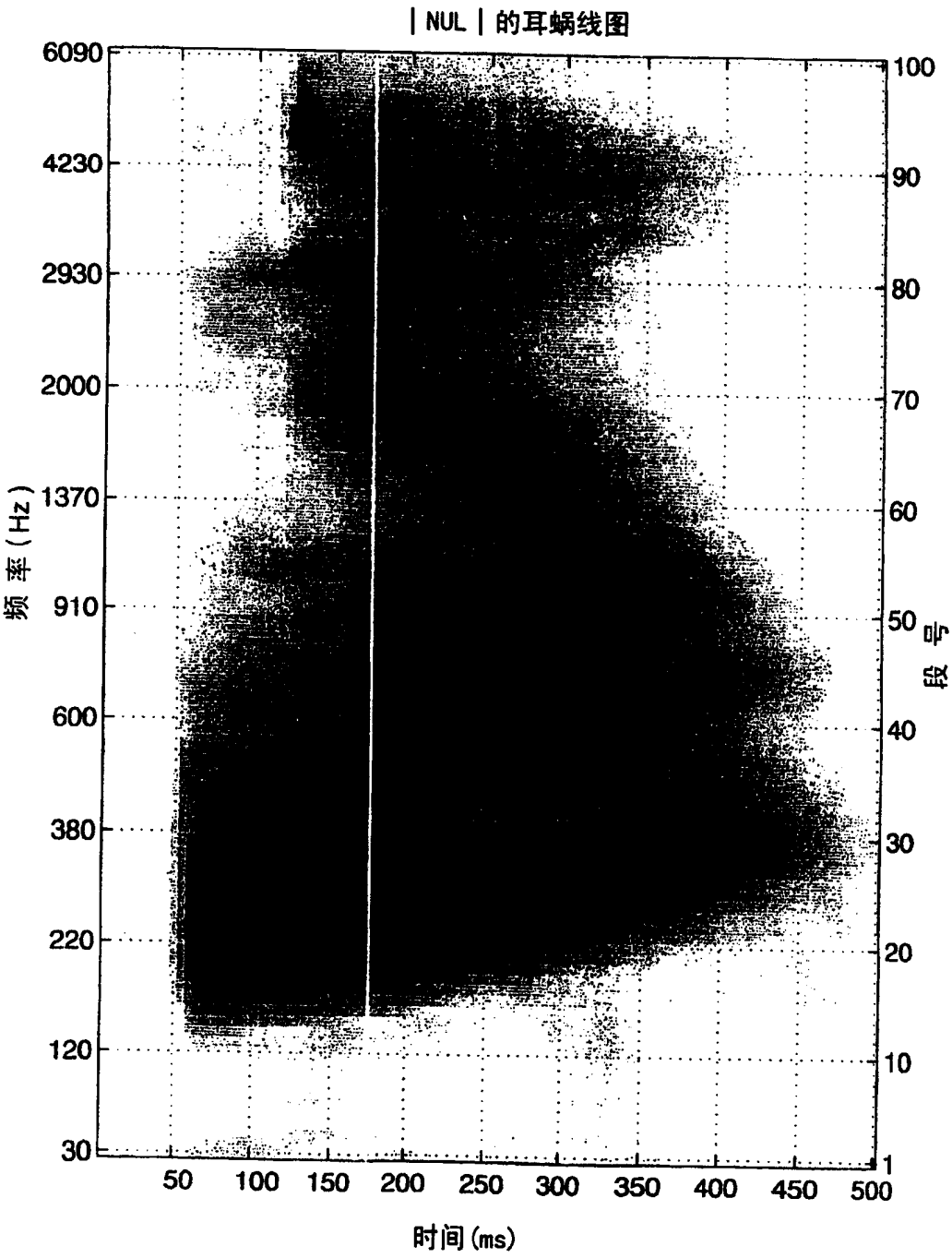


图2.2

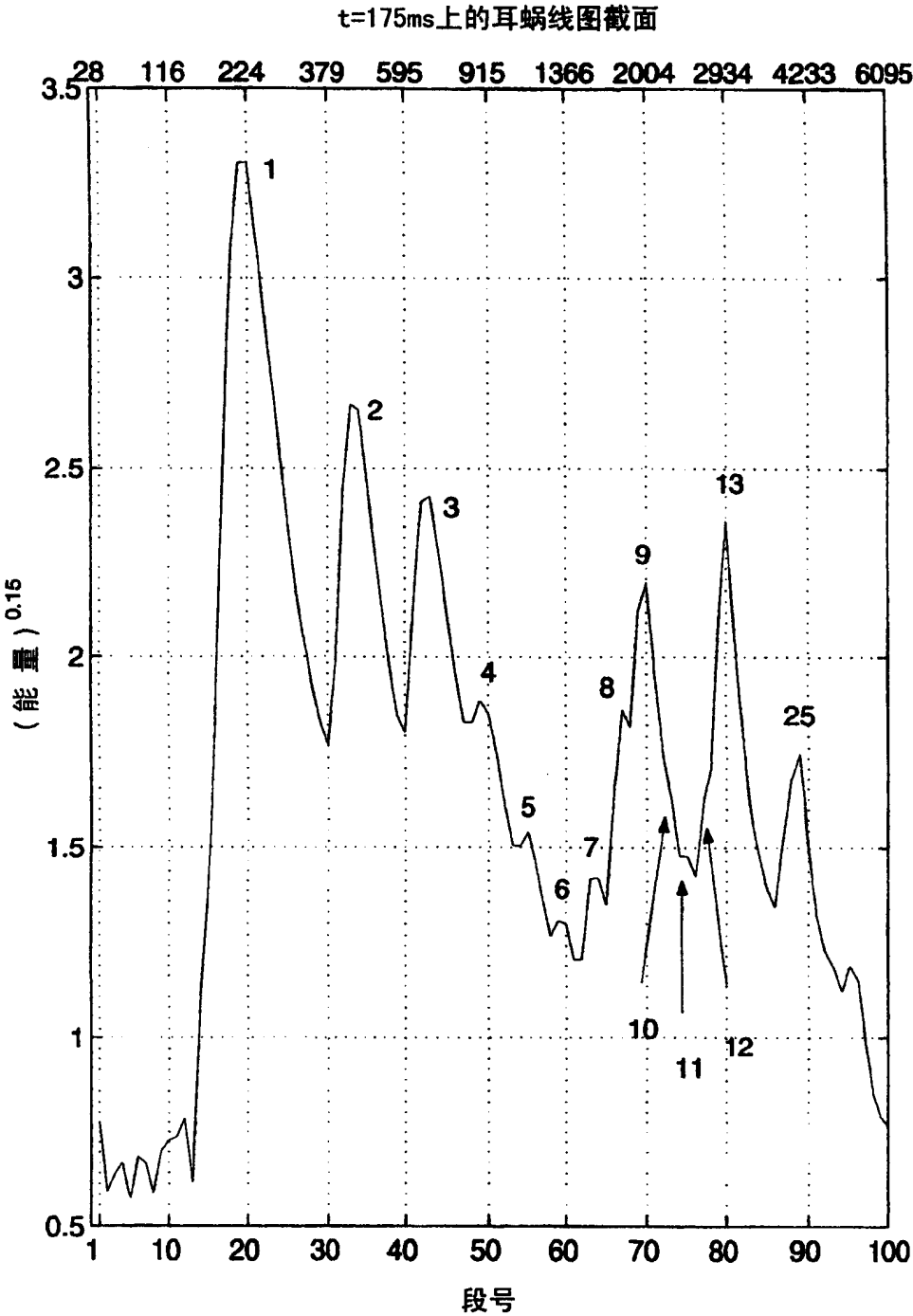


图2.3

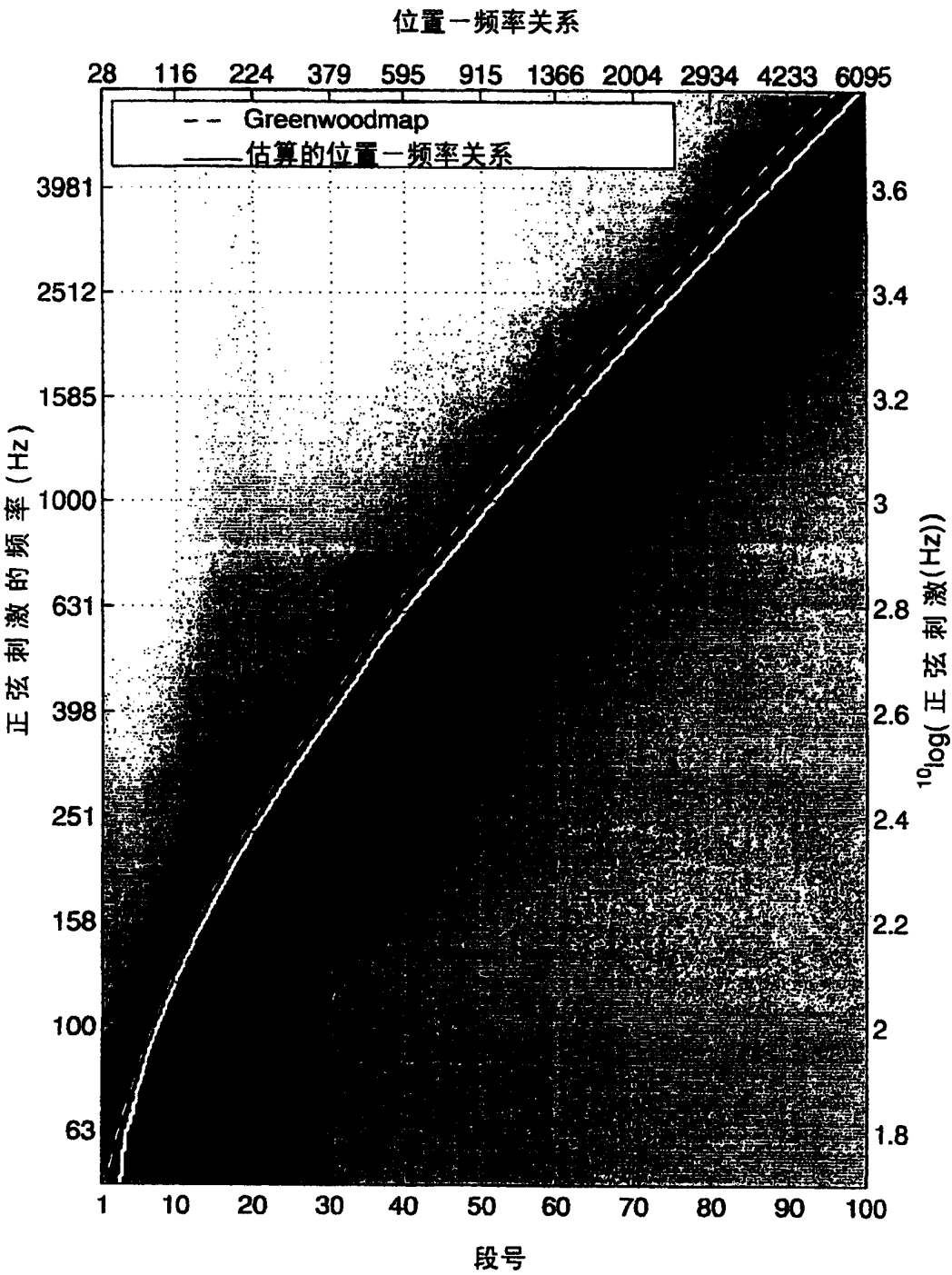


图2.4

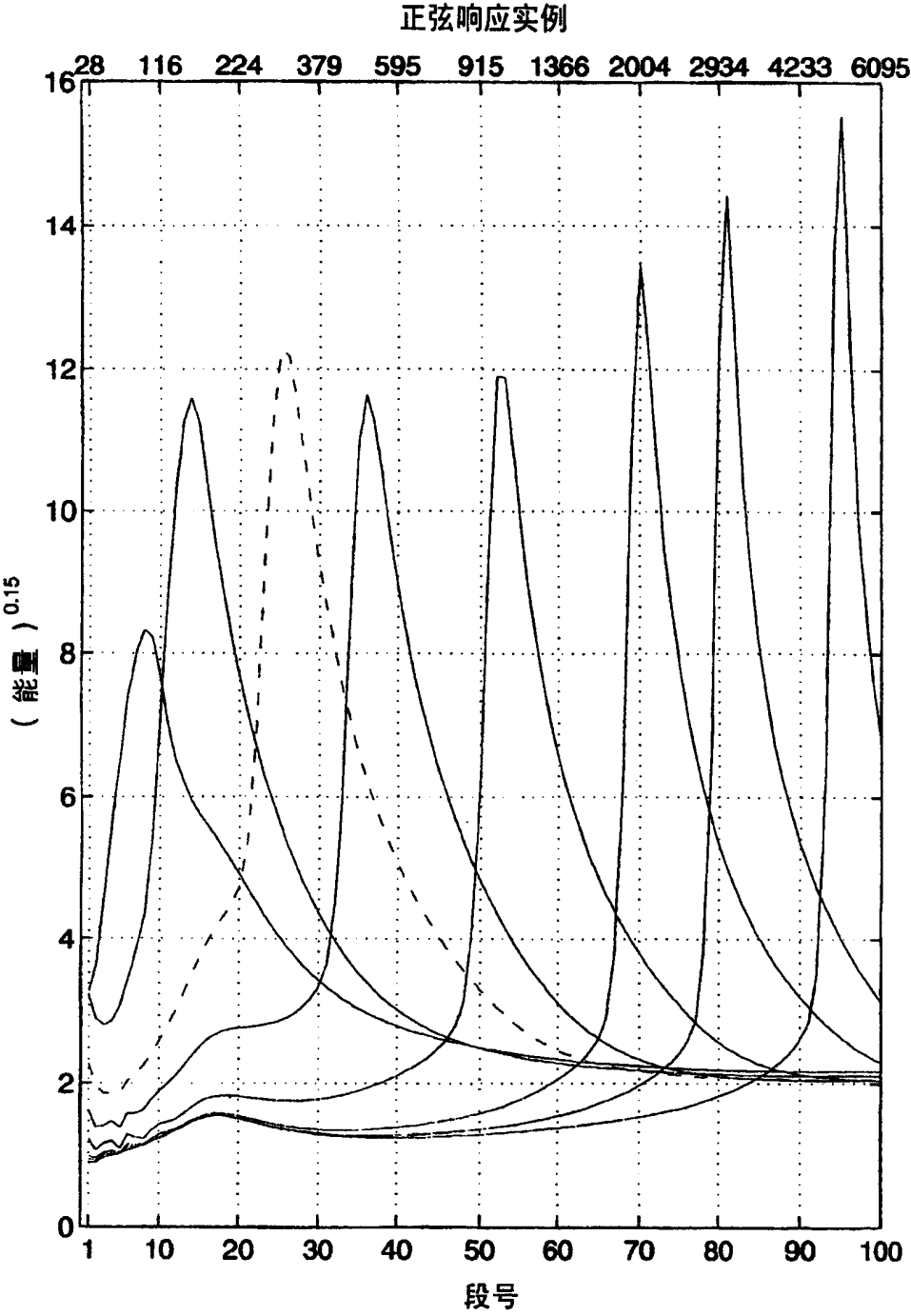


图2.5

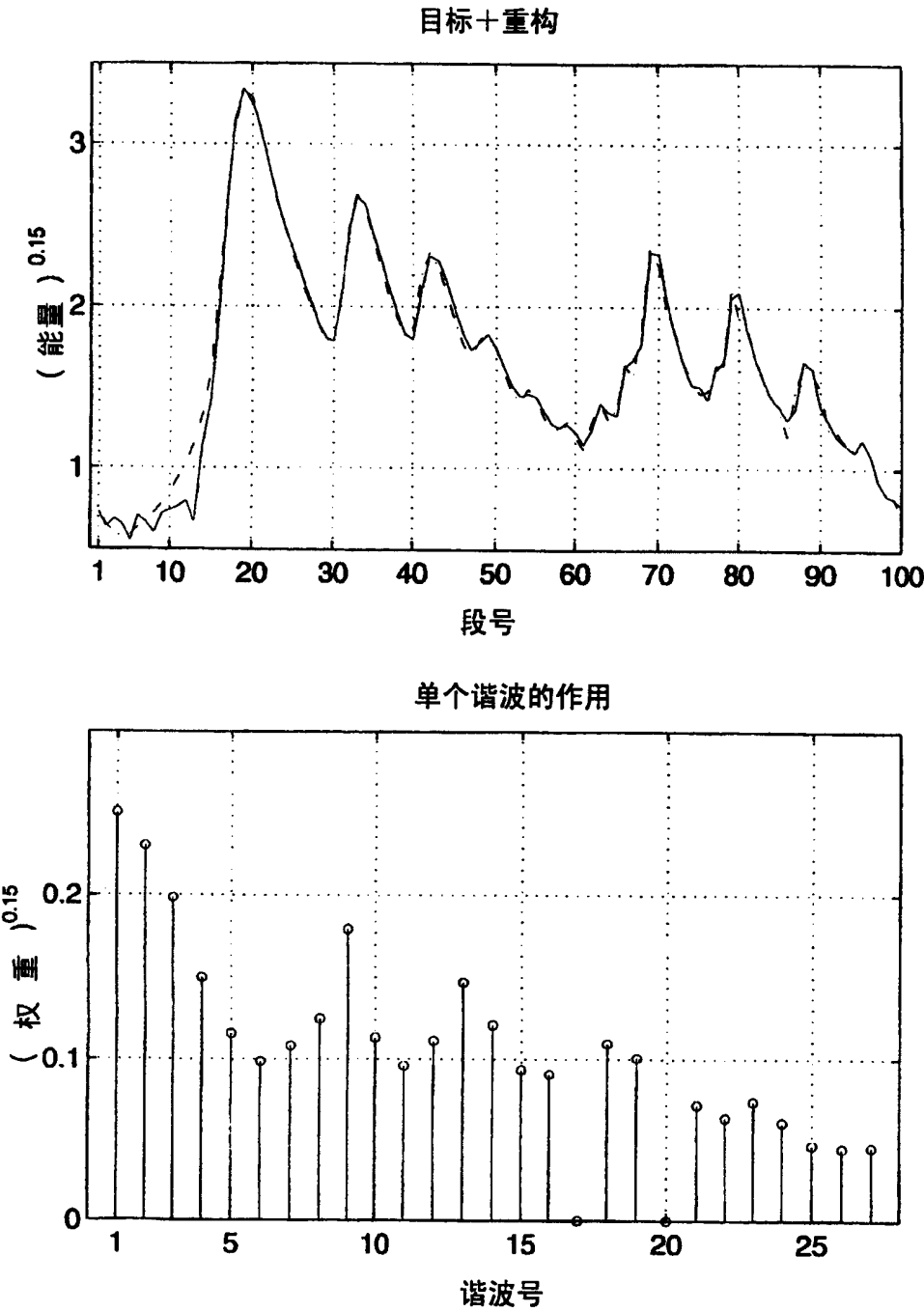


图2.6

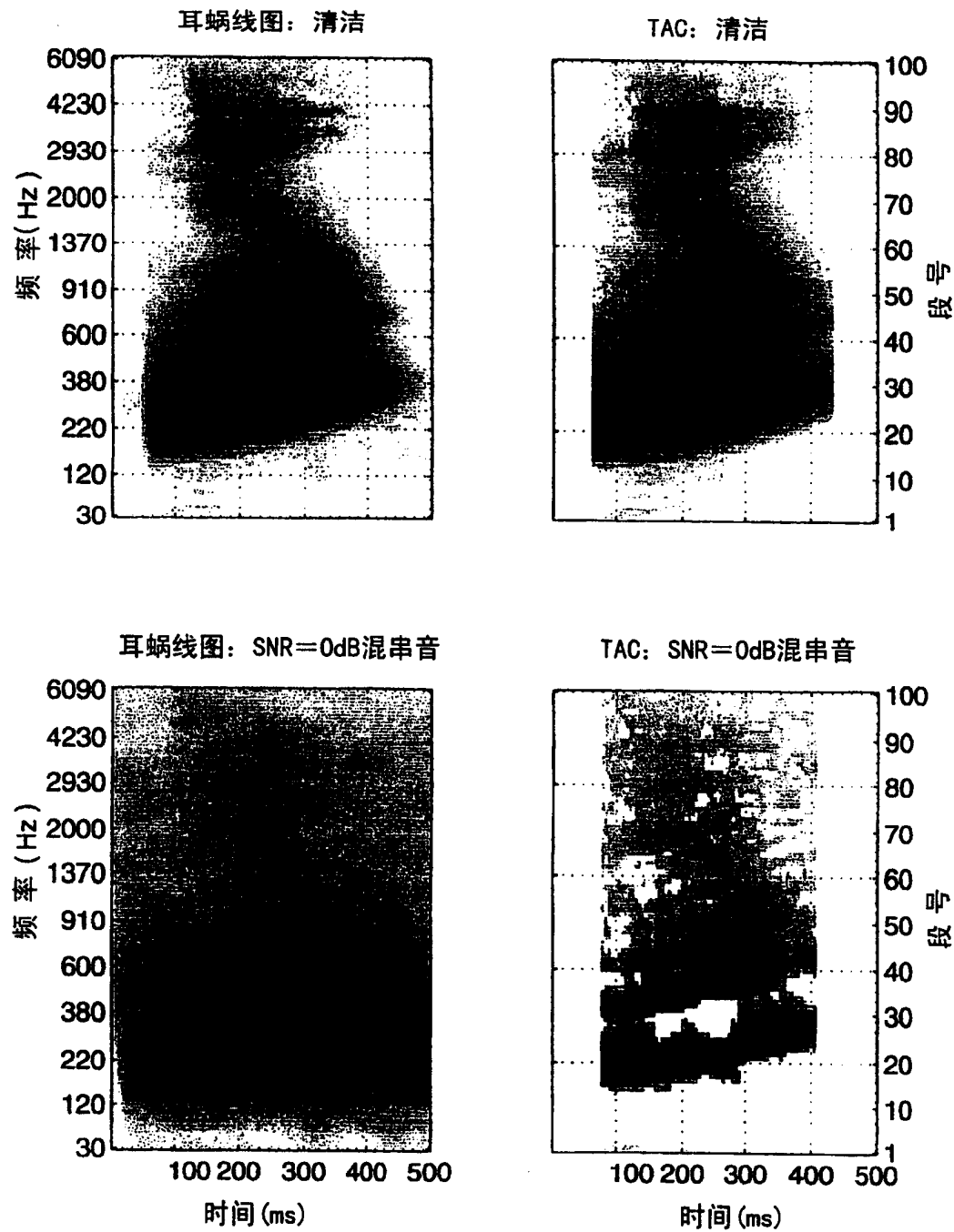


图2.7

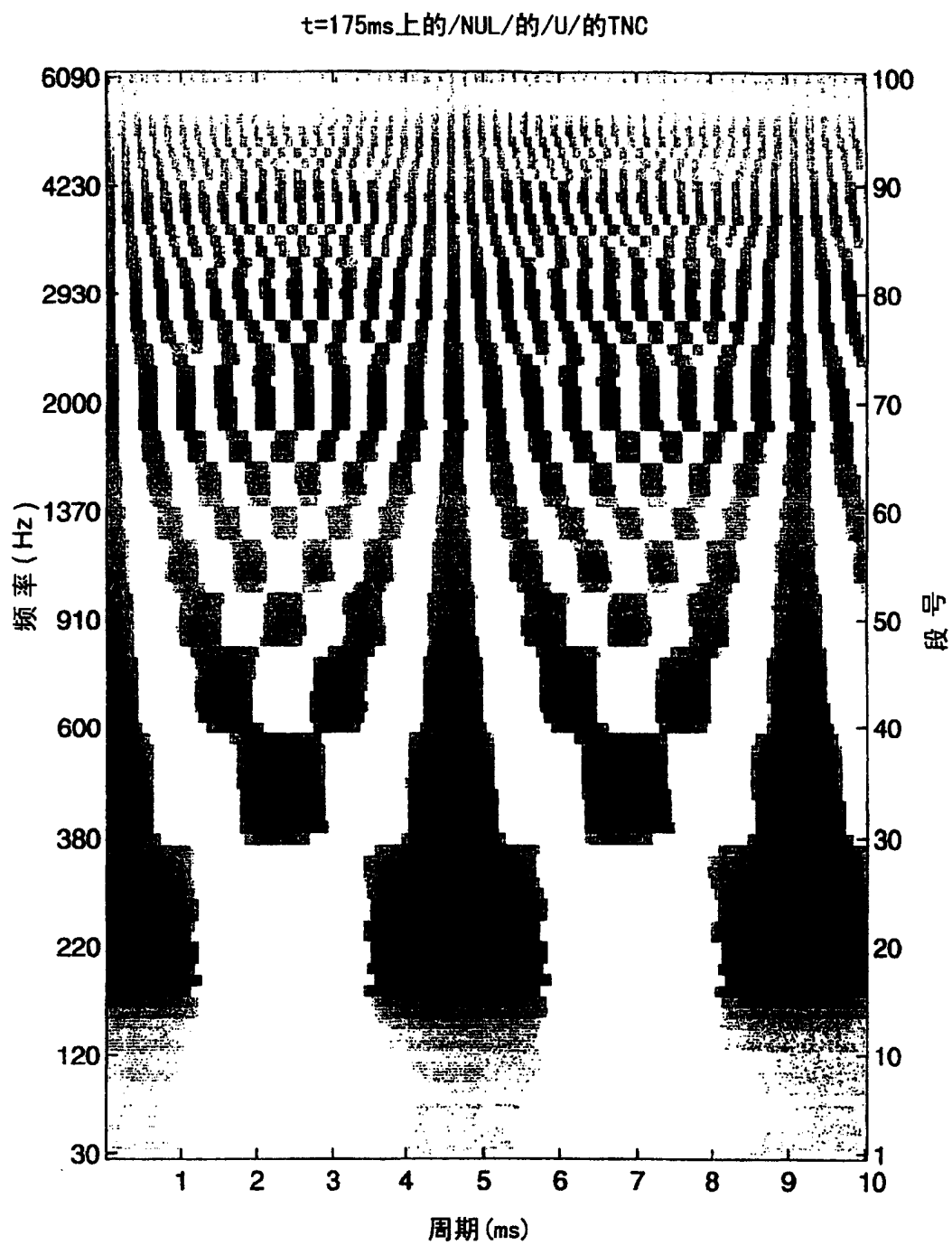


图 2. 8

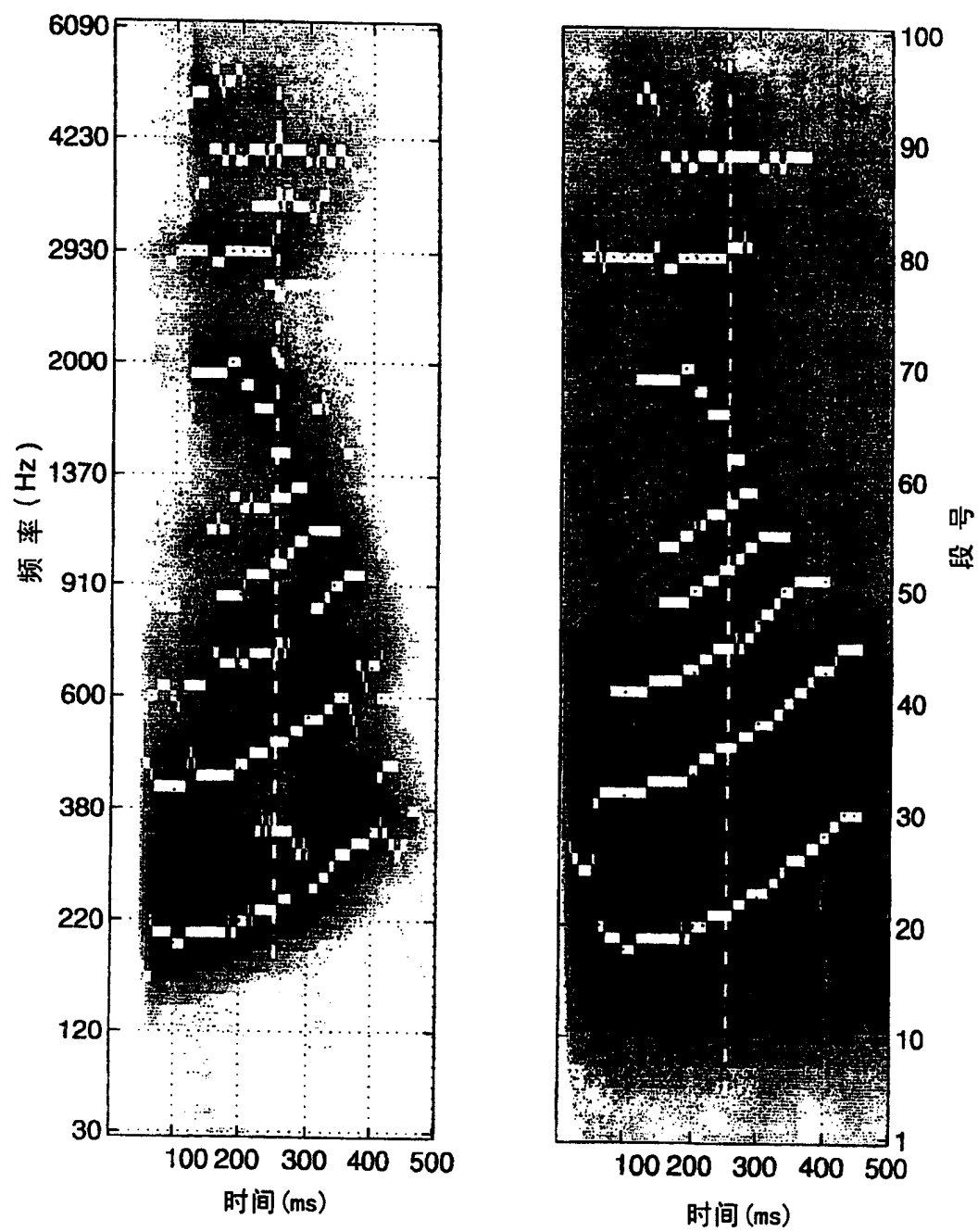


图2.9

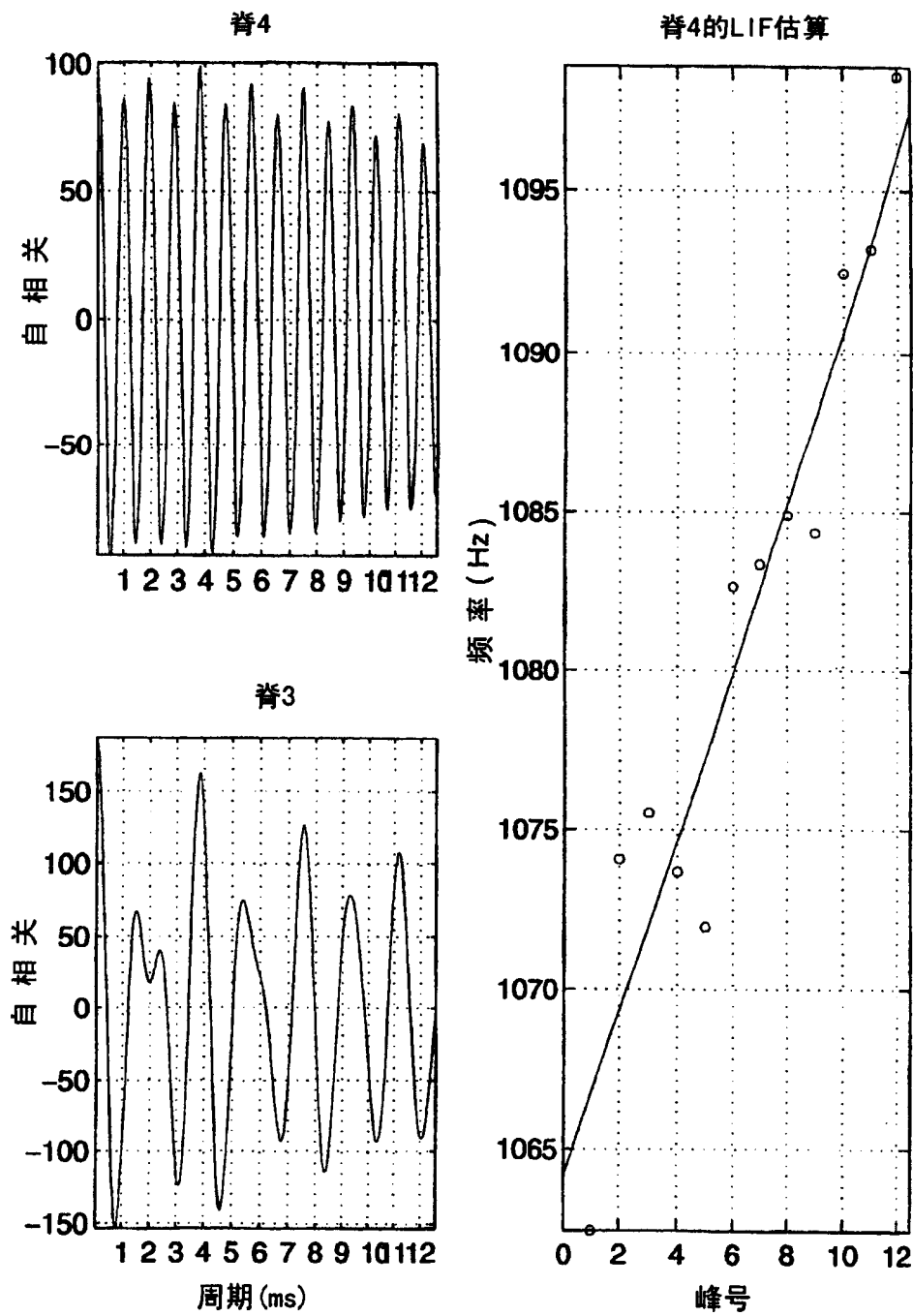


图2.10

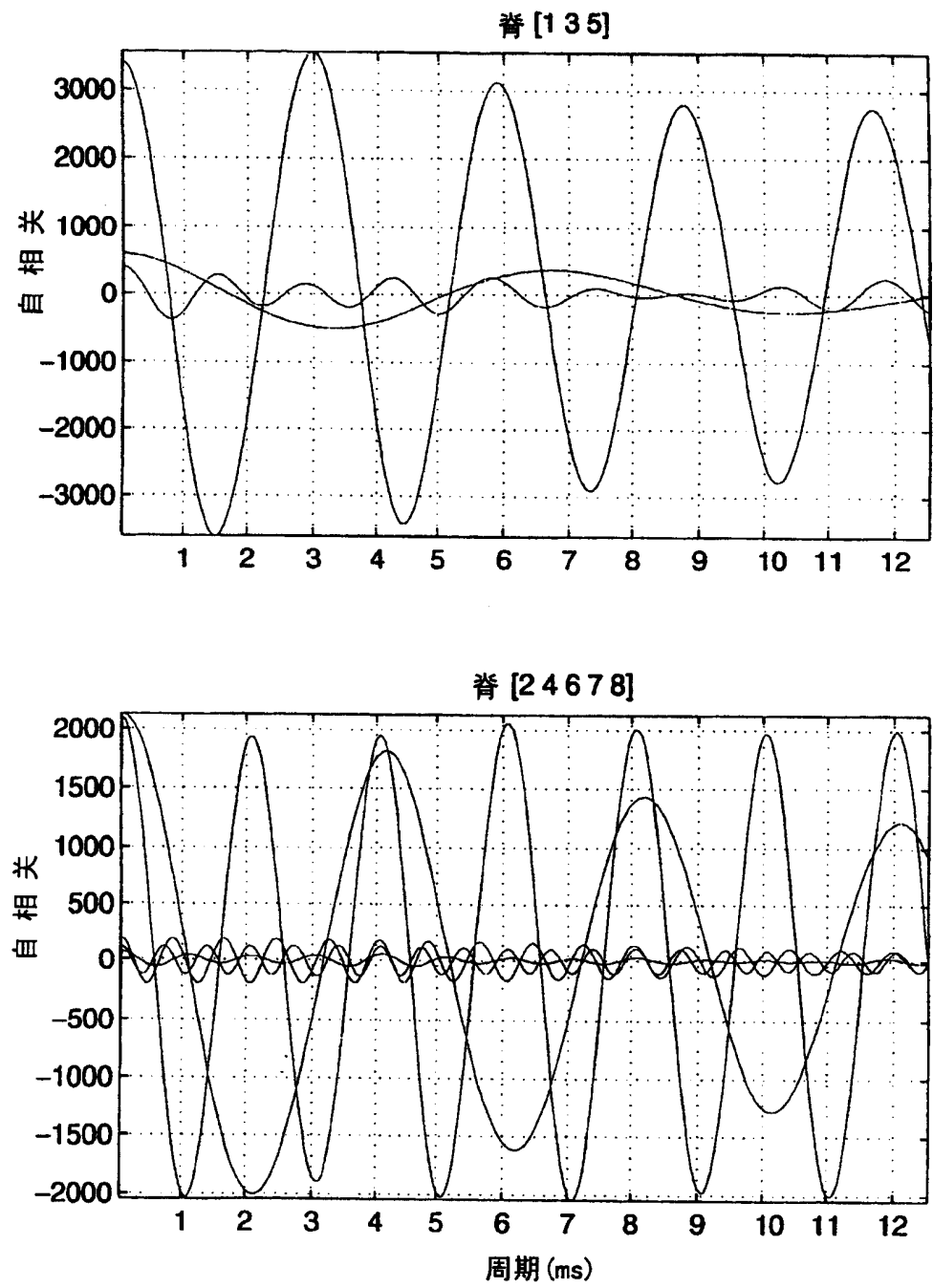


图2. 11

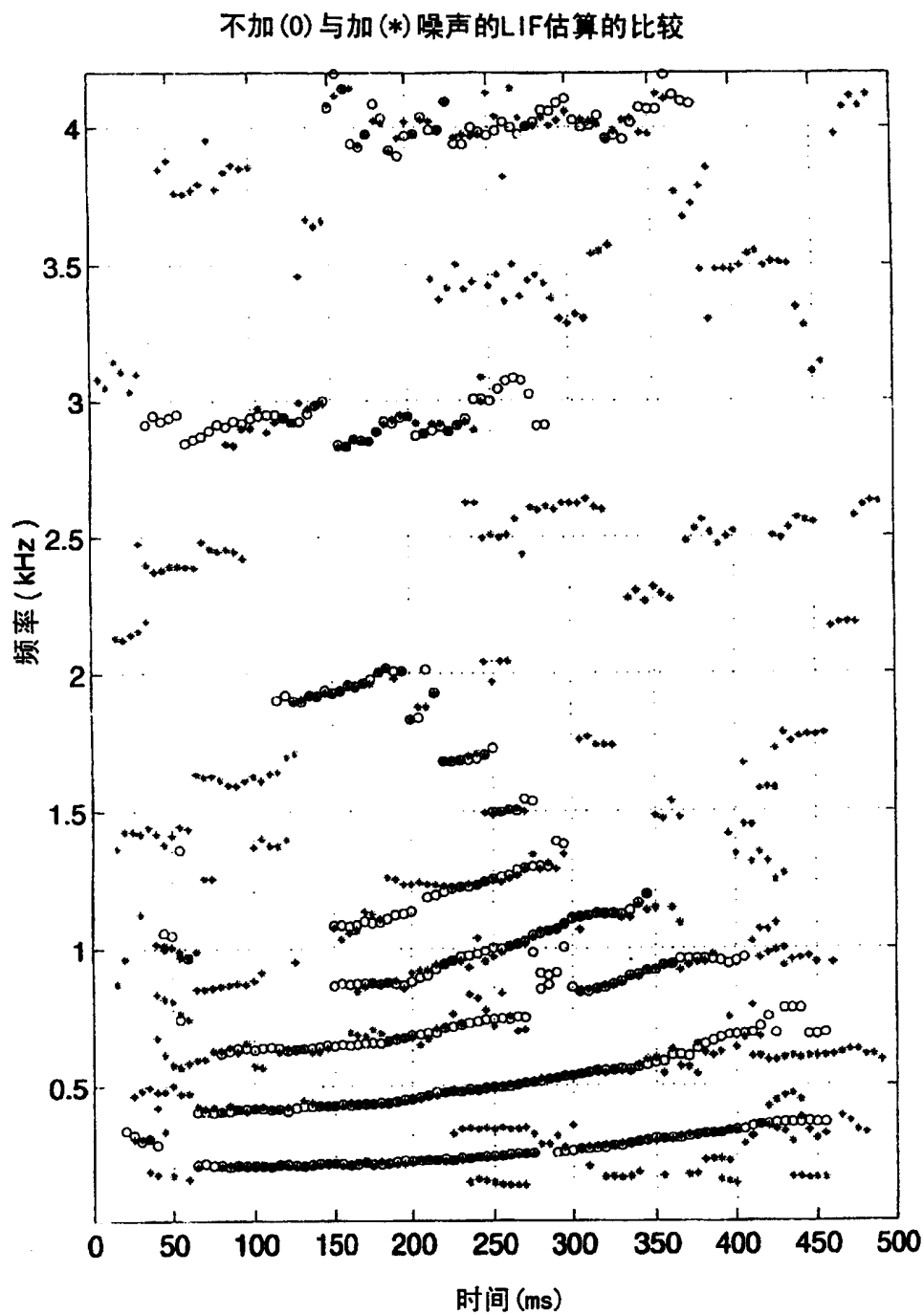


图2.12

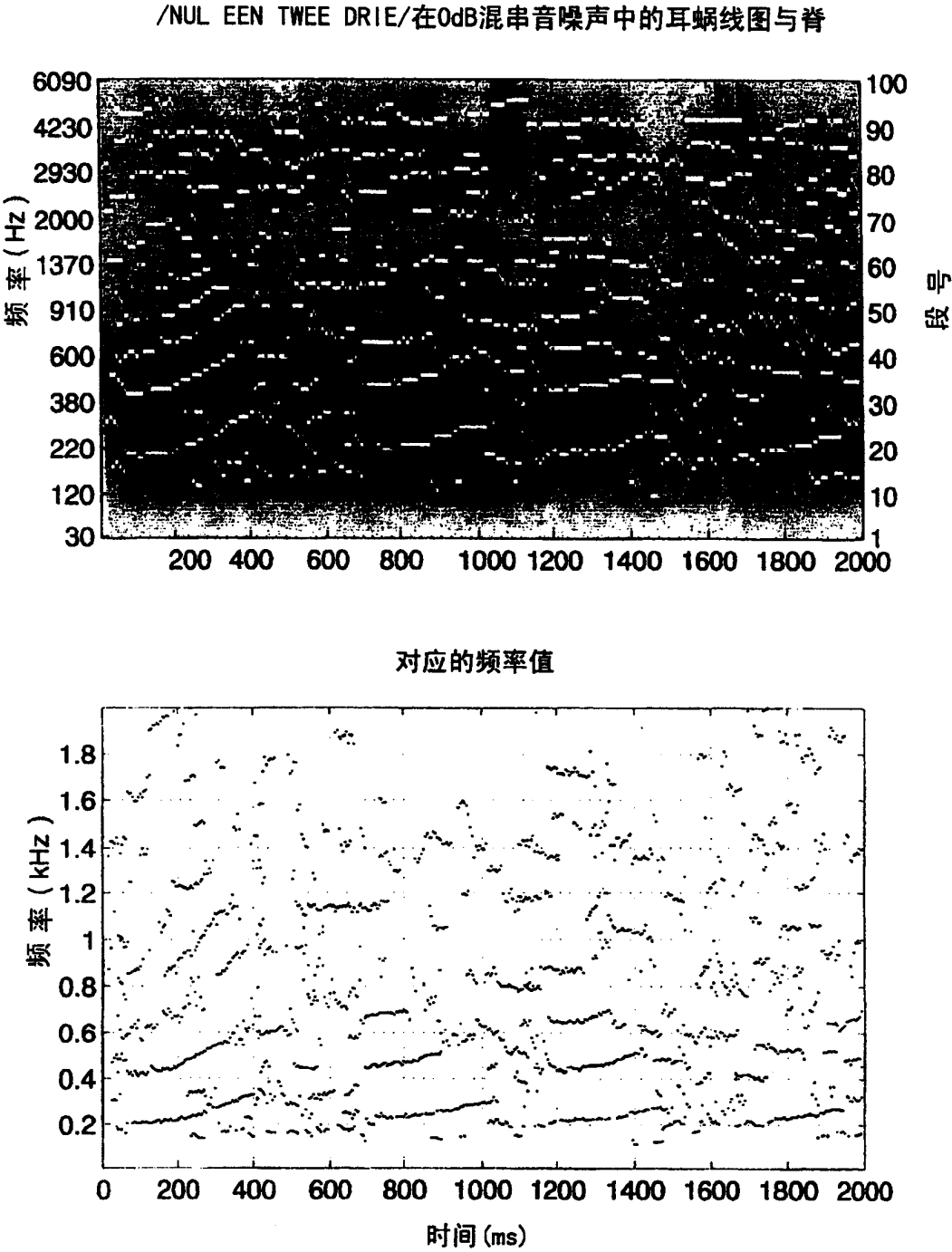


图2. 13

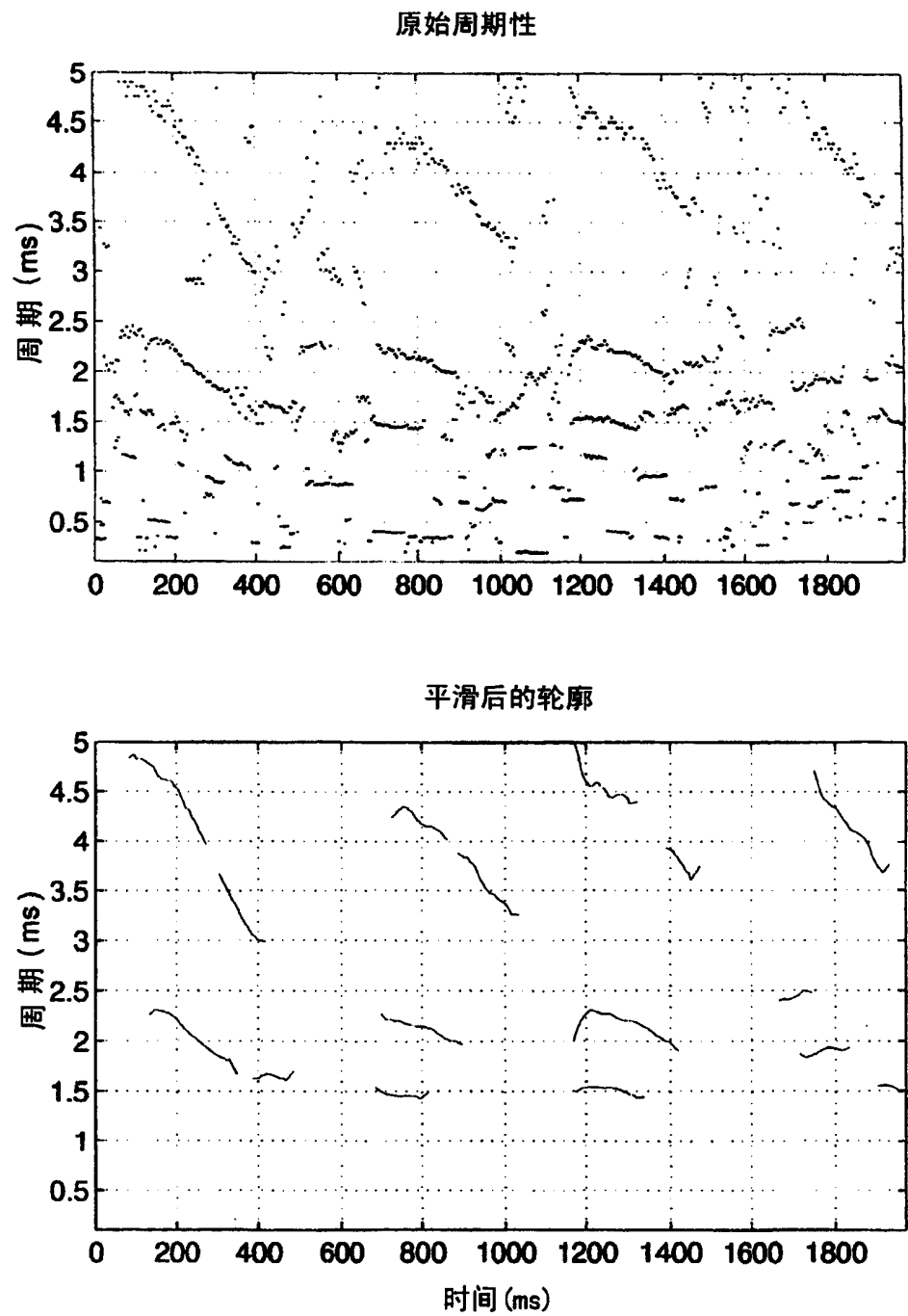


图2.14

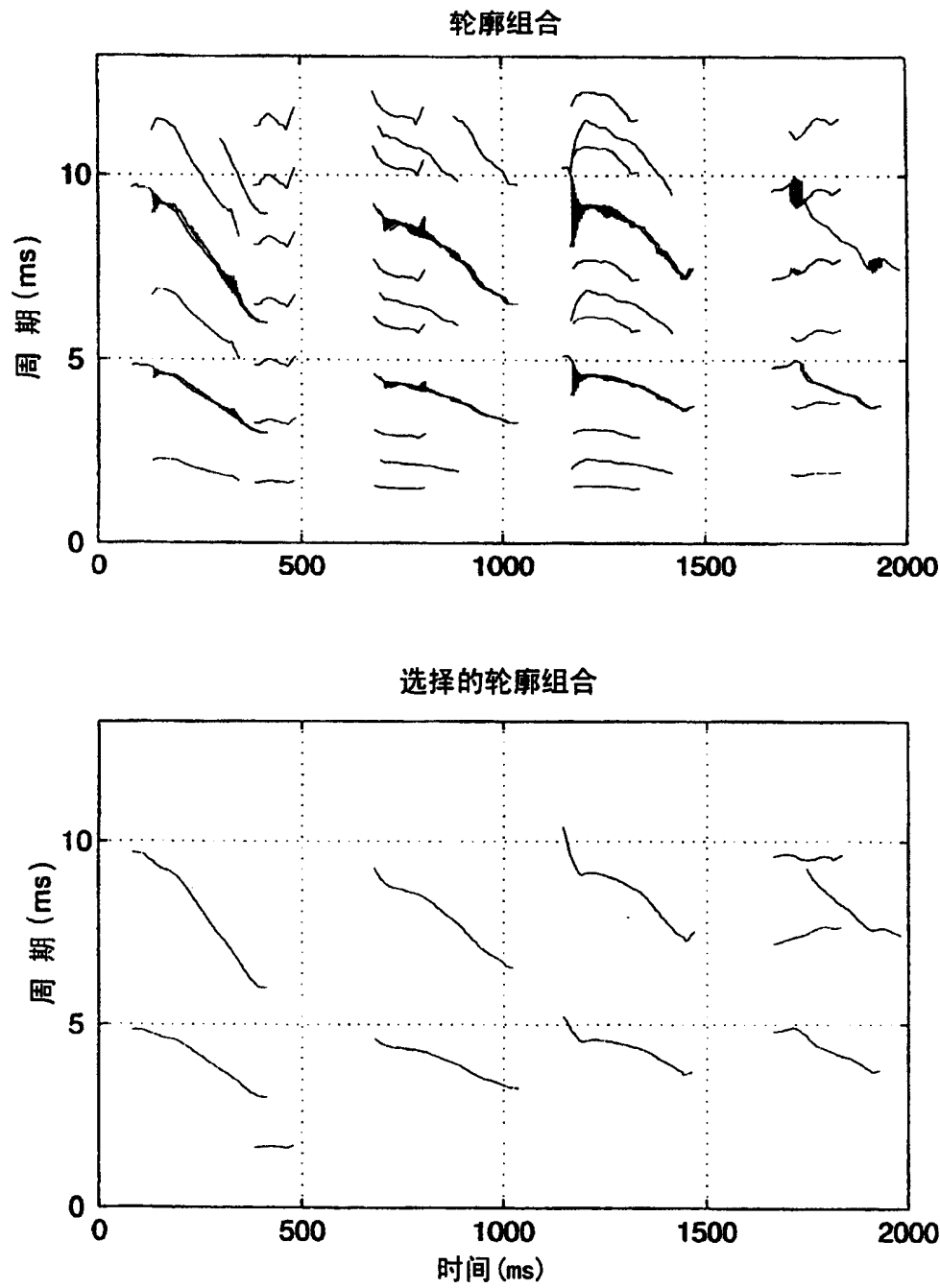


图2. 15

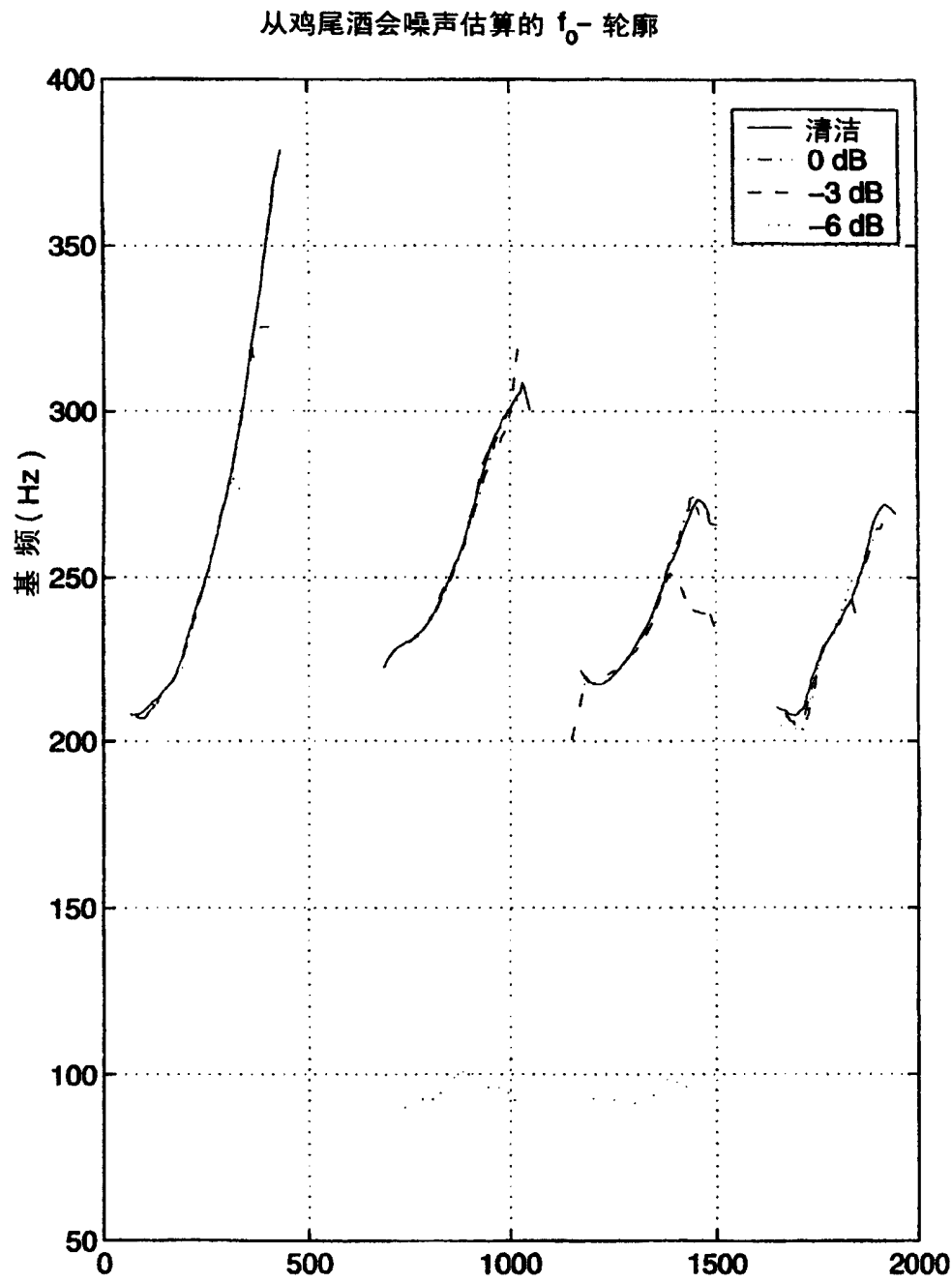


图2. 16

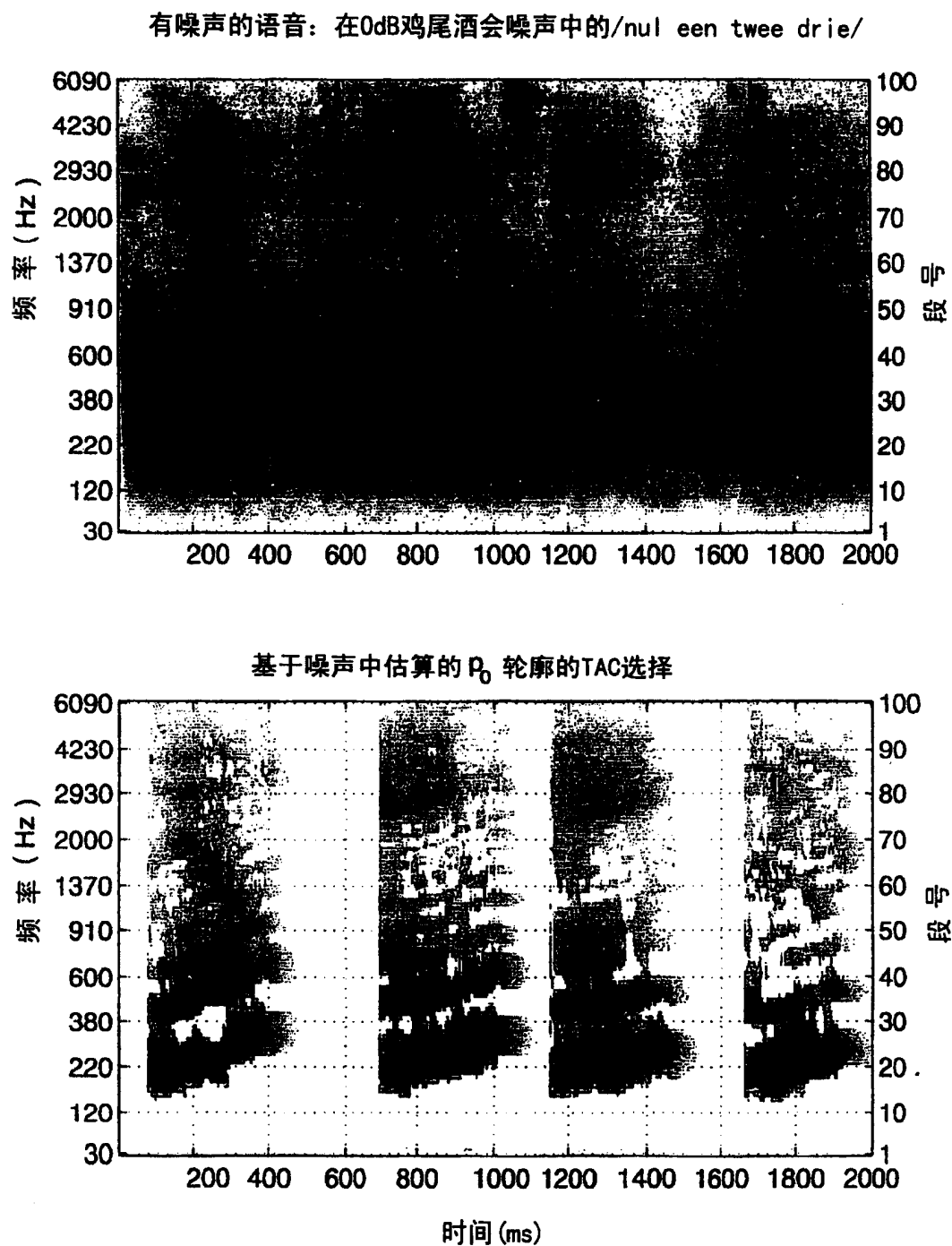


图2.17

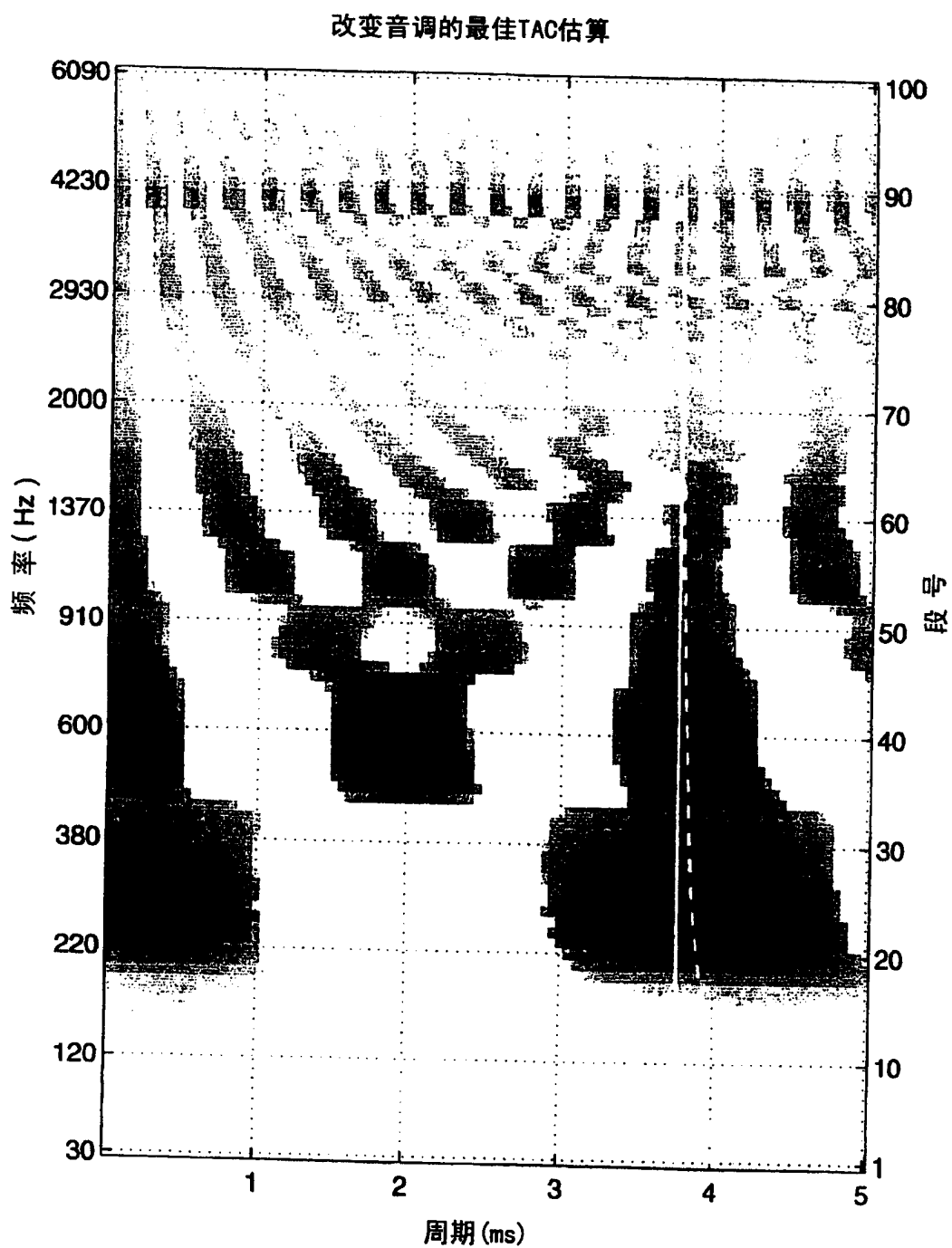


图2. 18

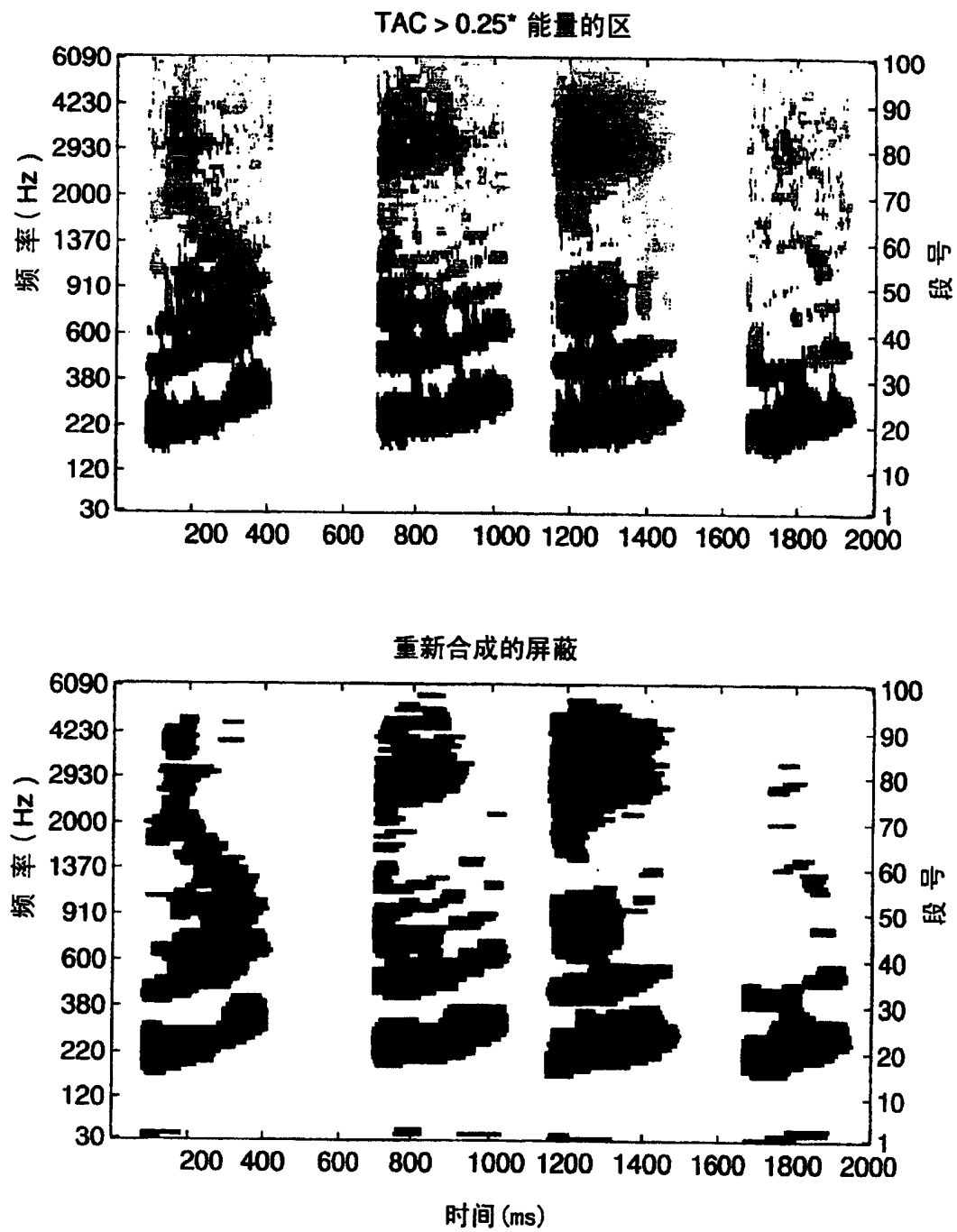


图 2. 19

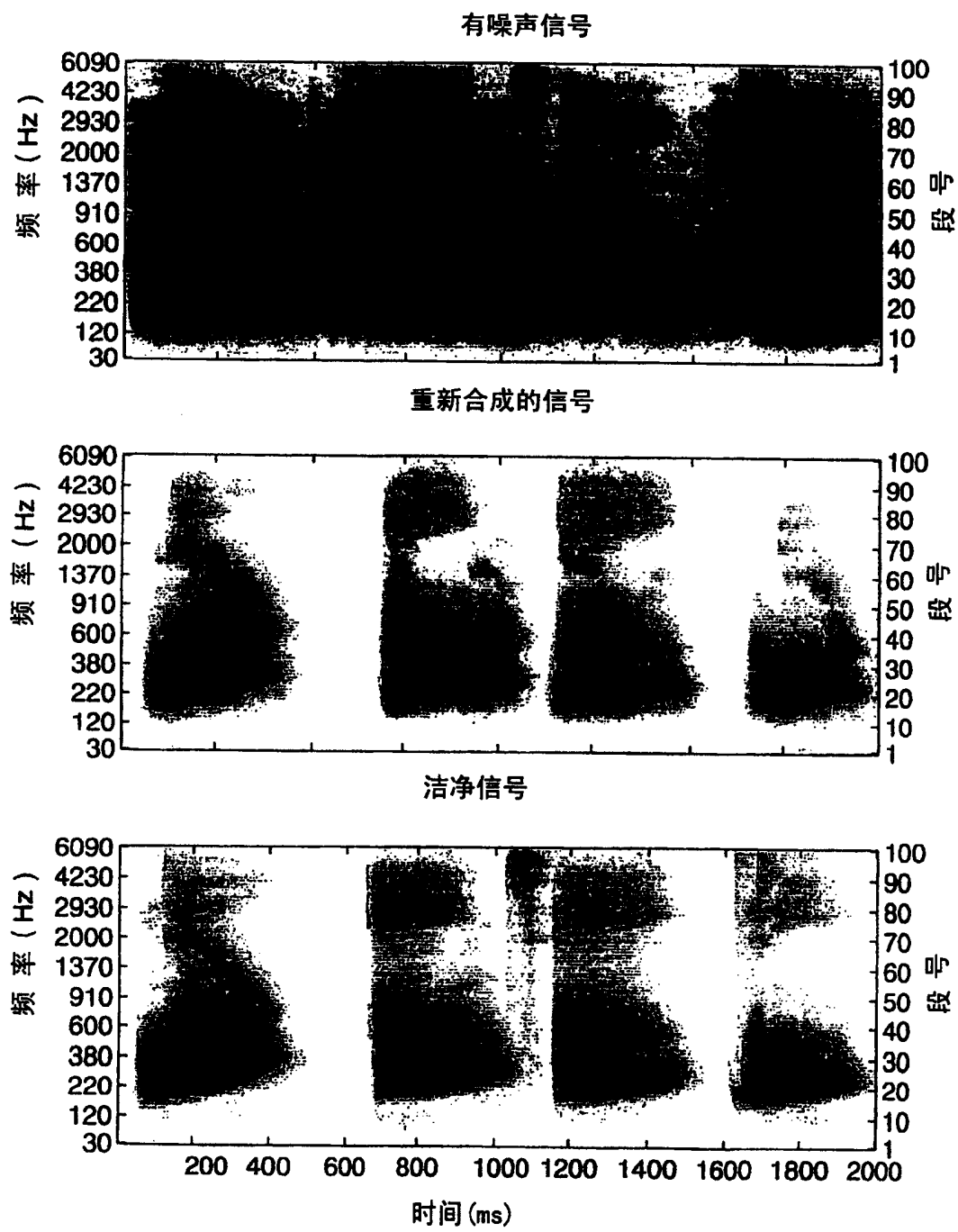


图2. 20

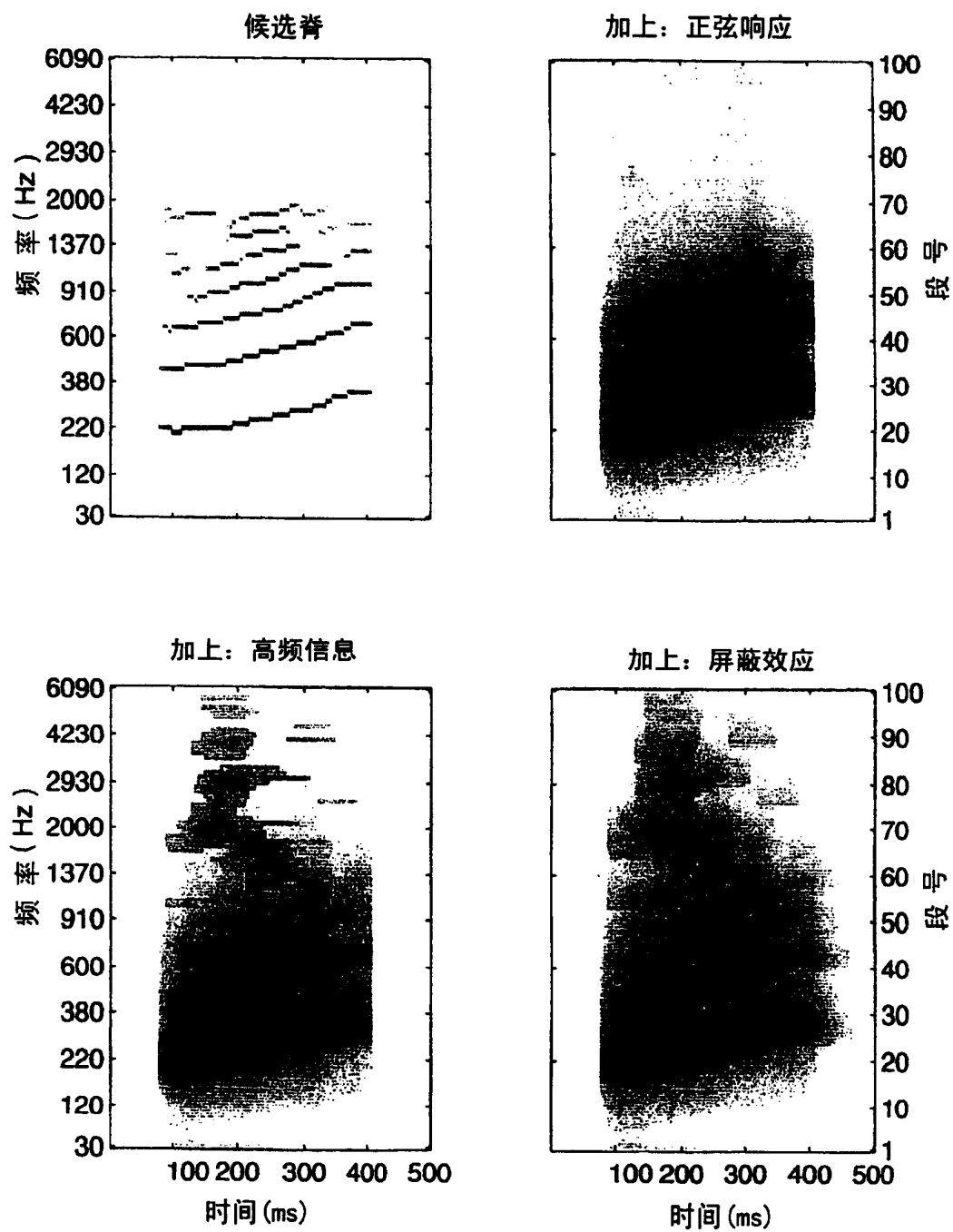


图2.21

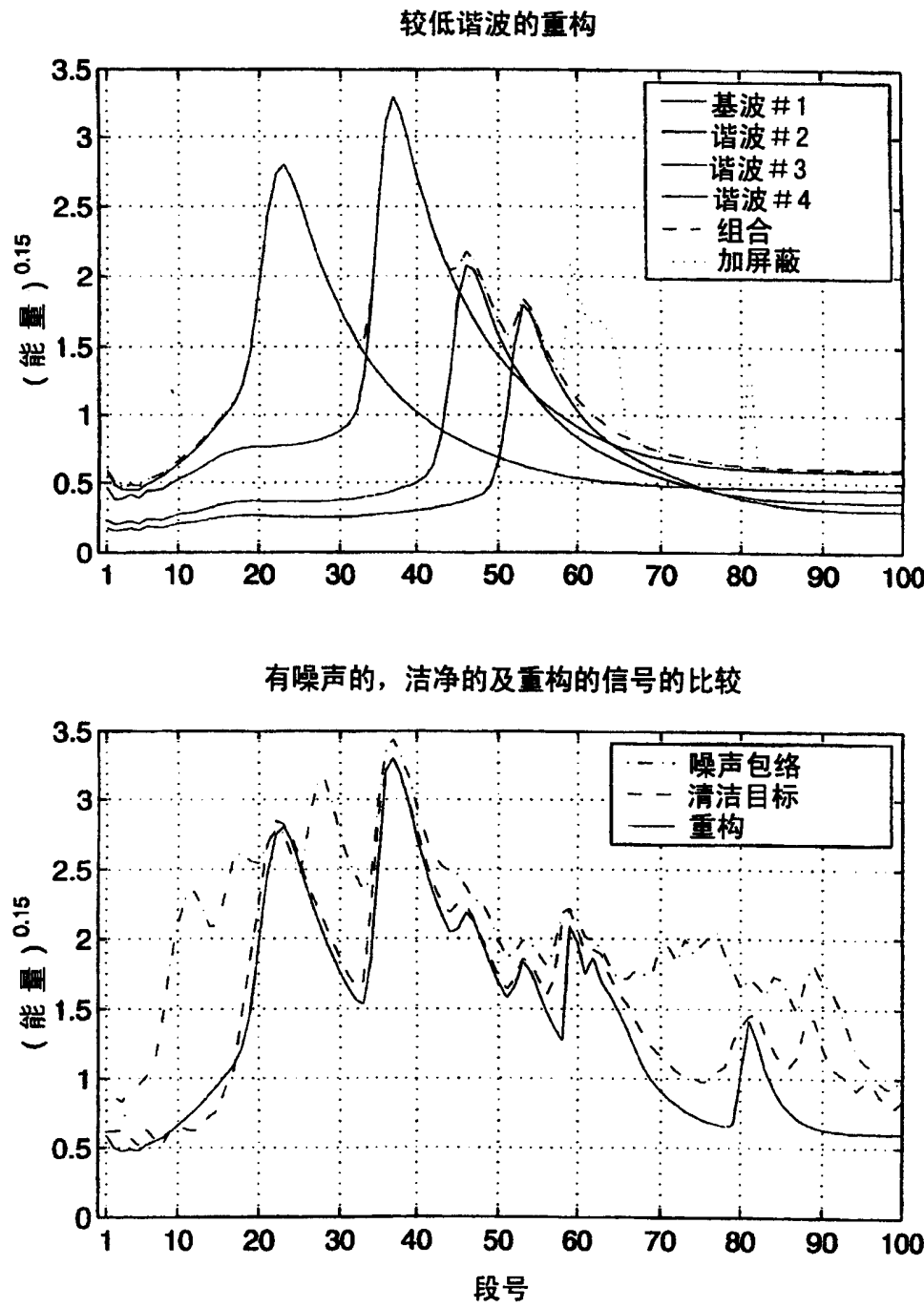


图2. 22

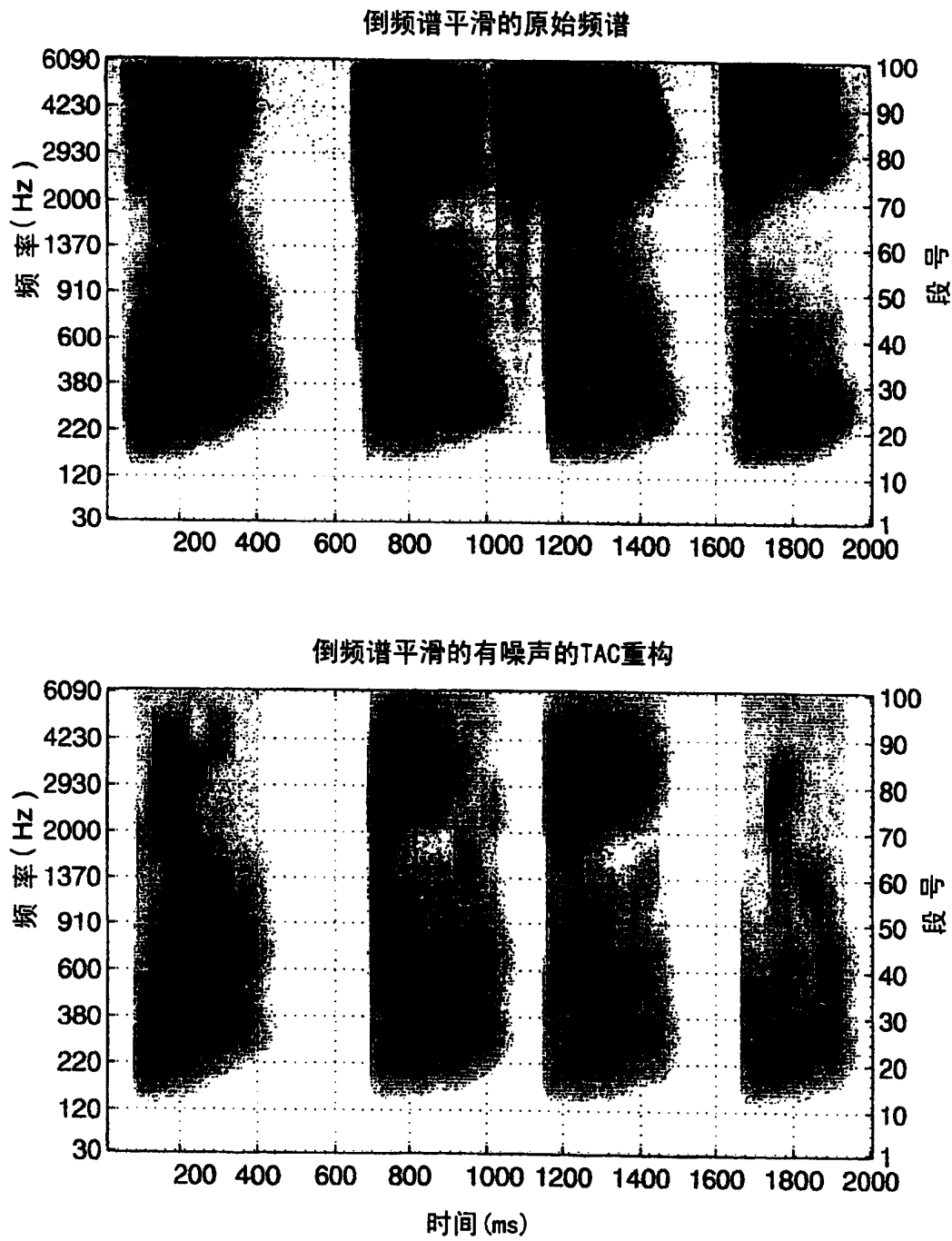


图2. 23

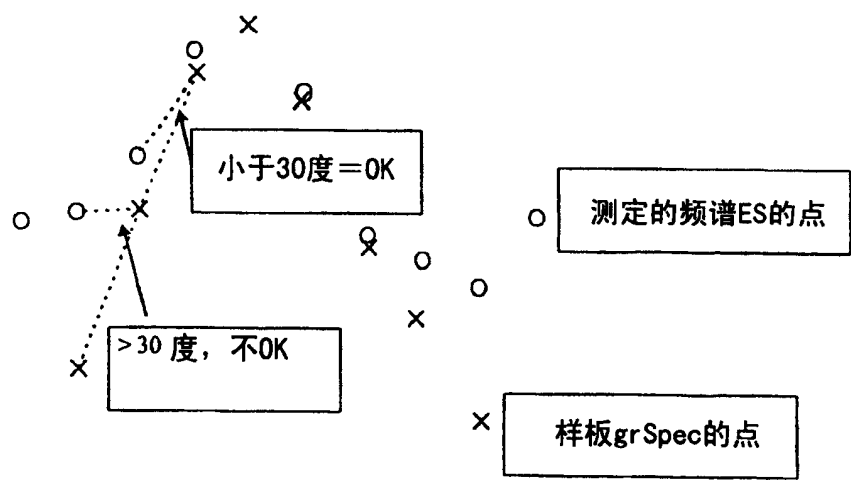


图3. 1

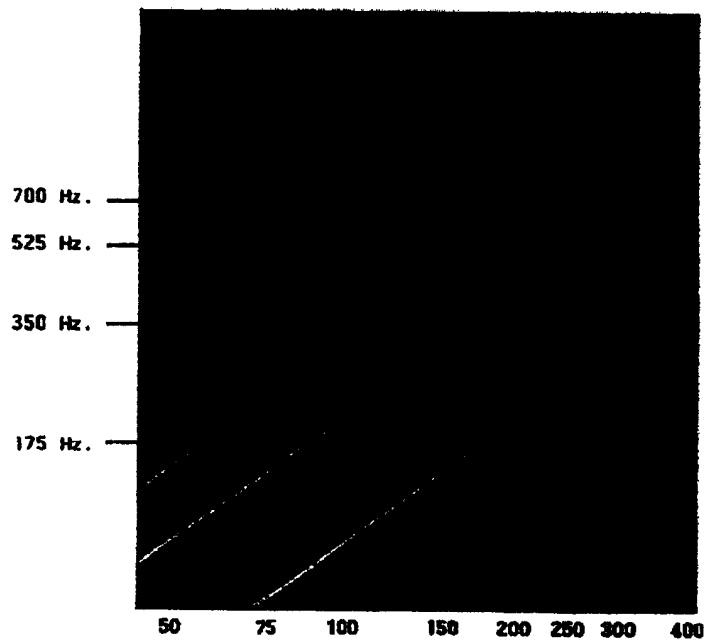


图3. 5A

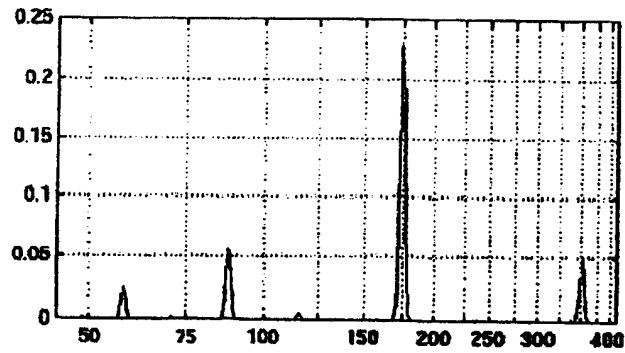


图3. 5B

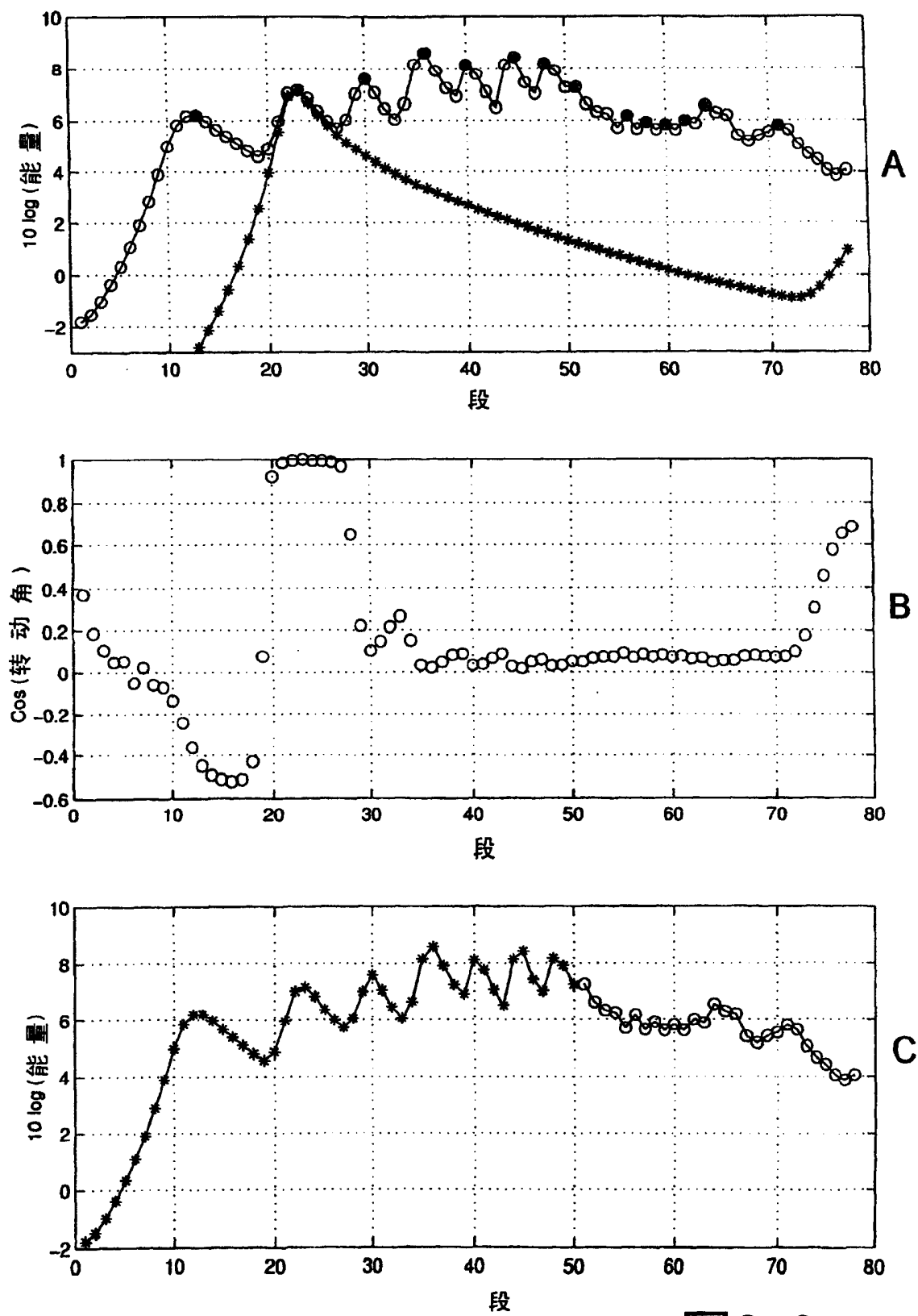


图3.2

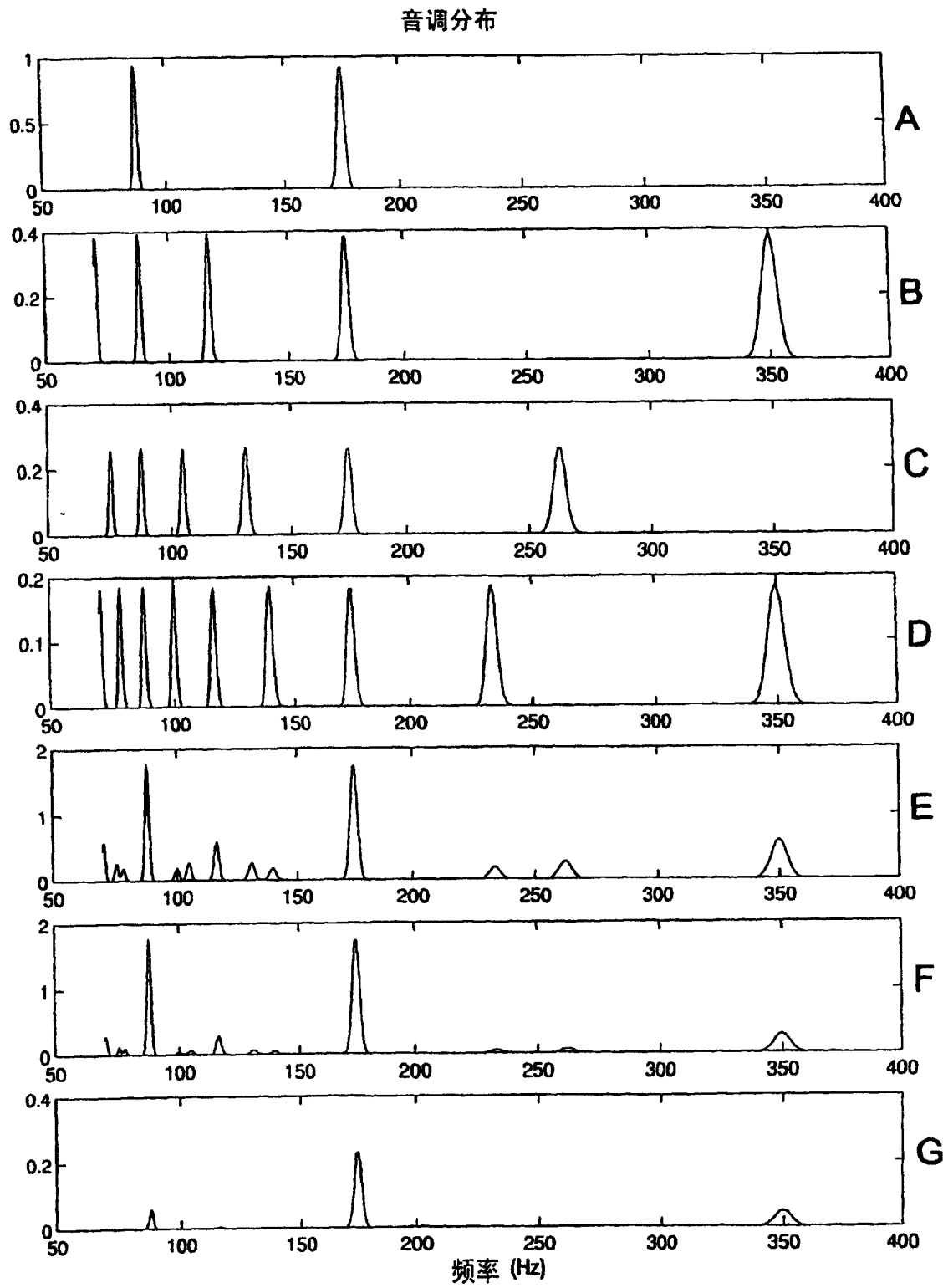


图3.3

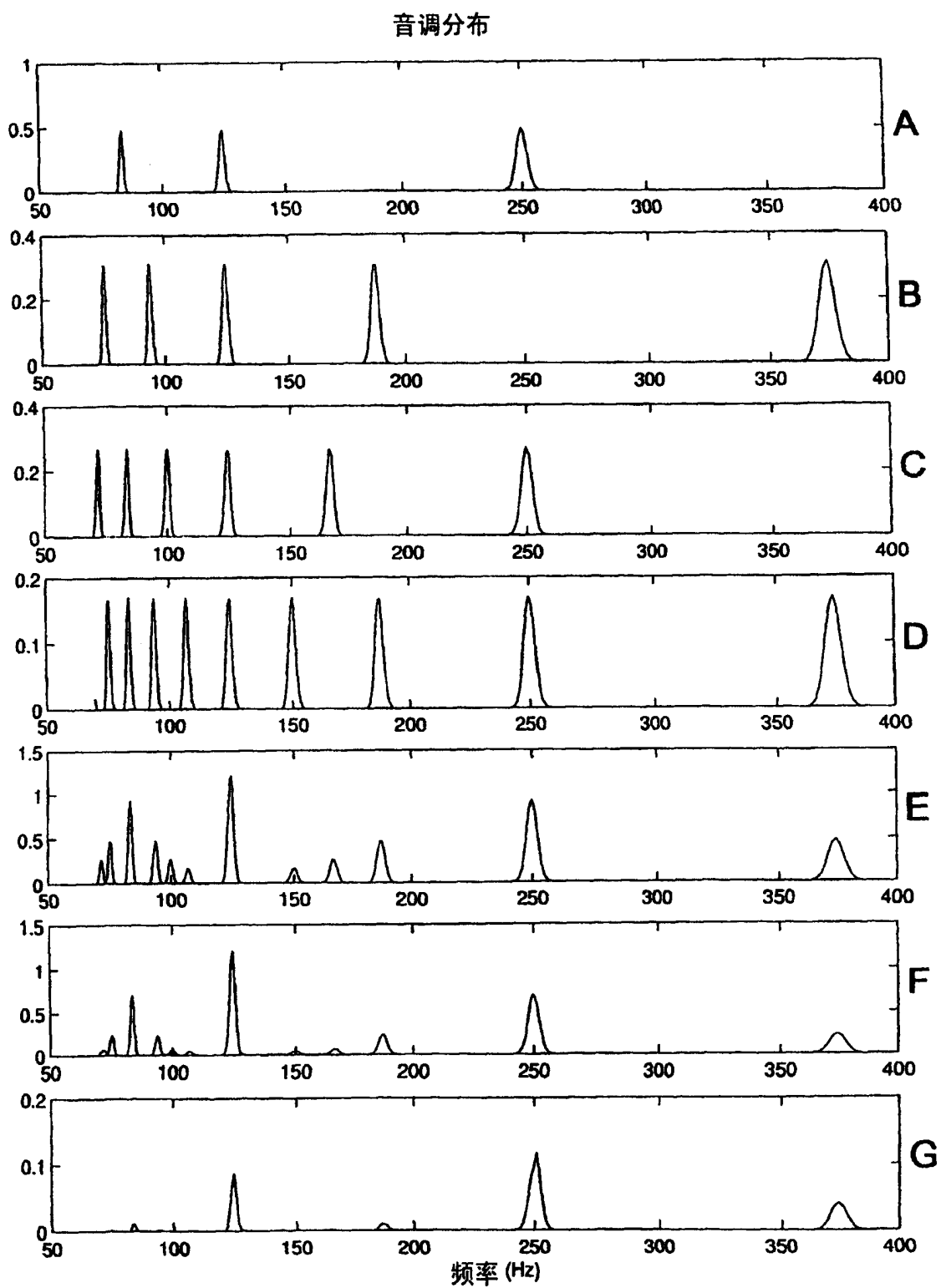


图3.4

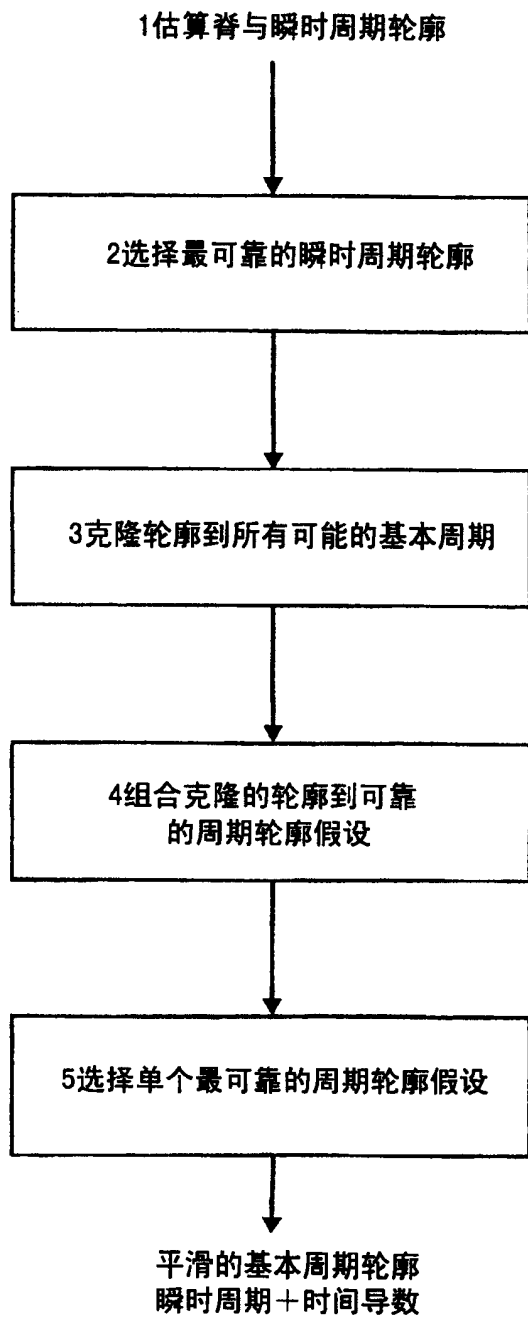


图3.6

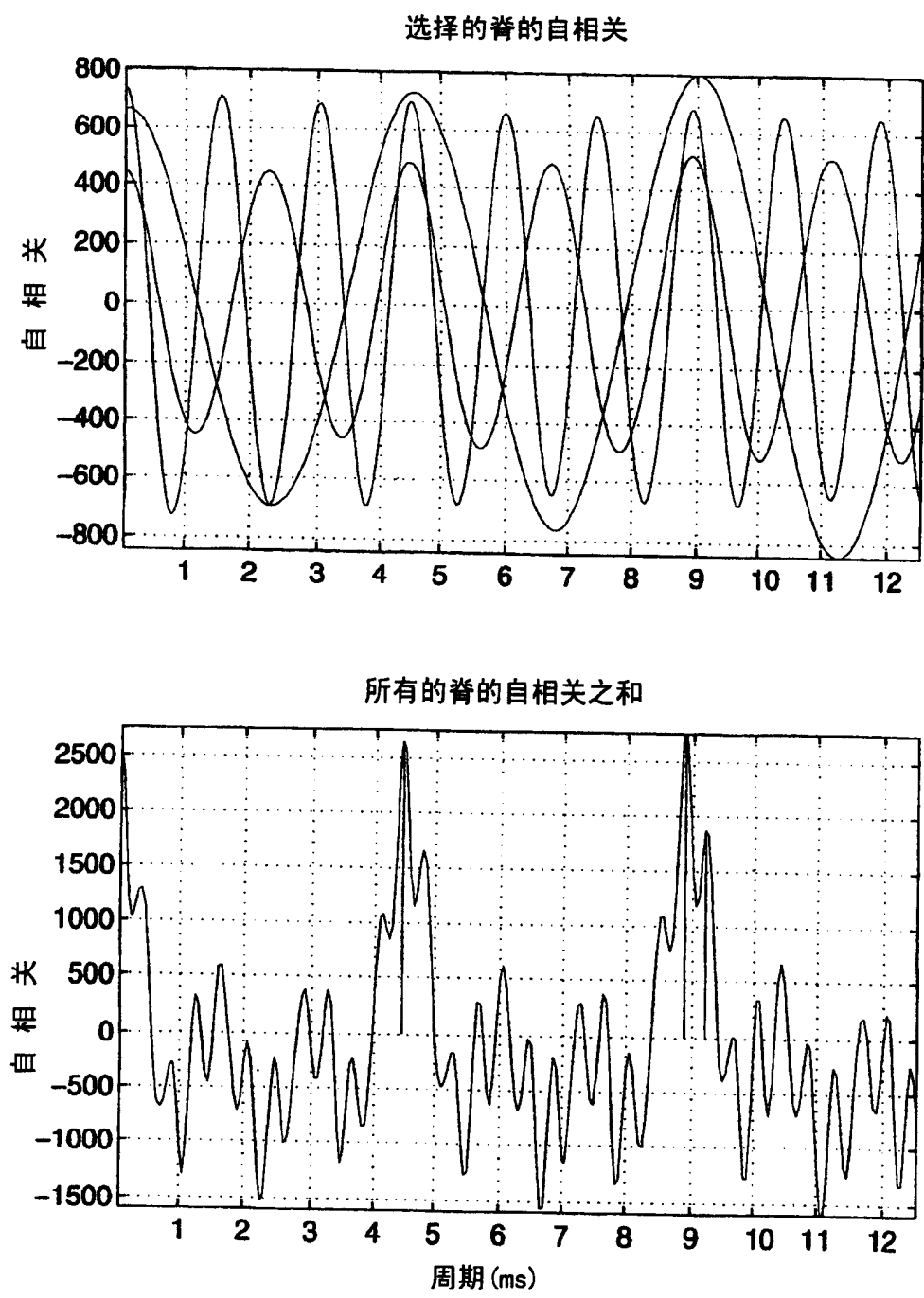


图3.7