

融合神经网络的卡尔曼滤波啸叫抑制路径突变检测算法

郭昊诚, 陈 锴, 卢 晶

(南京大学声学研究所, 近代声学教育部重点实验室, 南京 210093)

摘要: 分区频域卡尔曼滤波 (Partitioned block frequency domain Kalman filtering, PBFDKF) 因其收敛速度快、稳态误差小的优势被应用在自适应滤波声反馈抑制 (Adaptive feedback cancellation, AFC)。然而, 当声反馈路径发生突变时, 卡尔曼滤波会进入锁死状态, 难以再次跟踪。本文提出一种融合神经网络的卡尔曼滤波啸叫抑制状态检测算法 (Kalman-filter-based AFC with state detection model, KFSD)。该系统将卡尔曼滤波声反馈抑制系统的传声器采集信号、残差信号和滤波器更新量作为输入特征, 通过神经网络对卡尔曼滤波的状态误差协方差矩阵进行修正, 从而实现路径突变情况下的再次跟踪和收敛。仿真实验结果验证了所提算法具有较高的正判率、较低的虚警率和较短的延迟帧数, 算法同时具备快速再跟踪性能, 提高了声反馈抑制效果。

关键词: 声反馈抑制; 自适应滤波; 卡尔曼滤波; 状态检测; 深度神经网络

中图分类号: TP301 **文献标志码:** A

Kalman - Filter - Based Acoustic Feedback Cancellation with State Detection Model for Fast Recovery from Abrupt Path Changes

GUO Haocheng, CHEN Kai, LU Jing

(Key Laboratory of Modern Acoustics, Institute of Acoustics, Nanjing University, Nanjing 210093, China)

Abstract: The partitioned block frequency domain Kalman filter (PBFDKF) has been applied in acoustic feedback cancellation (AFC) due to its fast convergence and low steady-state misalignment. However, the Kalman filter at steady state might encounter the issue of deadlock when the feedback path experiences abrupt changes, exhibiting suboptimal tracking capabilities. In this paper, the Kalman-filter-based AFC with state detection model (KFSD) is proposed to effectively improve the robustness against abrupt path changes. The narrowband energy of the microphone signal, the residual signal and the update of Kalman filter are used as the input to the state detection model. And then, the state detection results are merged into the state estimation error covariance matrix of the Kalman filter, achieving better re-convergence performance against the abrupt path changes. Experimental results demonstrate the superior performance of the proposed KFSD algorithm, showcasing a high true positive rate, a low false alarm rate, and a short state detection latency. These advantages lead to faster re-convergence and enhanced acoustic feedback cancellation.

Key words: acoustic feedback cancellation; adaptive filtering; Kalman filtering; state detection; deep neural network

基金项目: 国家自然科学基金(12274221)。

收稿日期: 2024-08-08; **修订日期:** 2024-09-10

引言

声反馈现象广泛存在于扩音系统中,扬声器泄露的声音被传声器采集后,形成闭环回路自激振荡,影响放音增益与音质,甚至出现啸叫^[1]。自适应滤波声反馈抑制法(Adaptive feedback cancellation, AFC)通过匹配声反馈路径将反馈声抵消,具有良好的啸叫抑制效果^[2-3]。对于公共放音系统或助听器场景,声反馈路径被拟合为一个高阶的自适应滤波器,为实现更短延时和更快运算,可使用分区频域自适应滤波算法^[4]。引入状态空间模型^[5-6],将声反馈路径用一阶马尔可夫模型表示,声反馈过程对应状态的观测方程,可以将自适应声反馈抑制与卡尔曼滤波相结合,即可得到分区频域卡尔曼滤波算法(Partitioned block frequency domain Kalman filtering, PBFDFK)^[7-8]。该算法理论上对应最优的变步长,在收敛过程中有较大的更新步长,从而实现了快速收敛;进入稳态后有较小的更新步长,保证稳态下误差较小,从而得到了较好的音质。

然而对于声反馈路径发生突变的情况,卡尔曼滤波会进入锁死状态。尽管路径发生变化,需要再次匹配,但更新步长仍停留在一个较小的值,难以再次跟踪,导致啸叫重现^[9]。为此,需要在系统中引入状态检测模块,一旦检测到路径突变,就对卡尔曼滤波算法中的变量进行修正,进而实现再跟踪收敛。文献[10]提出了引入阴影滤波器的突变检测算法,该算法的主滤波器基于卡尔曼滤波,在与主滤波器平行的位置上设置了具有较大更新步长的阴影滤波器。主滤波器的卡尔曼滤波在正常工作状态下,与阴影滤波器相比具有更快的收敛速度与更小的匹配误差。发生路径突变时,阴影滤波器的更新步长较大,可快速追踪到新的声反馈路径,匹配误差较小;而主滤波器处于锁死状态,更新步长很小,再收敛速度很慢,造成较大的匹配误差。因此,可以通过比较主滤波器和阴影滤波器的宽带误差信号能量值来检测声反馈路径的状态突变。当主滤波器误差能量与阴影滤波器误差能量的比值大于某一阈值时,判断其发生了路径突变。该方法可以在一定程度上改善系统的跟踪性能,但因其未充分利用原有主滤波器结构中的参量特征,导致检测速度较慢,甚至会发生误判。

近年来深度神经网络(Deep neural network, DNN)被越来越广泛地应用到与音频声学相关的分类问题中,如语音端点检测^[11]、声学事件检测^[12]和啸叫检测^[13]等。受此启发,可以利用DNN在分类问题上的能力,将路径突变状态检测问题与其结合。在之前的相关工作中^[14],作者采取AFC-PBFDFK算法中的反馈损失增强比作为特征量,结合深度神经网络,进行路径突变的状态检测。本文在此工作的基础上进一步扩展,将传声器采集信号、残差信号和滤波器更新量作为特征,输入轻量级神经网络用于状态检测,对卡尔曼滤波的状态误差协方差矩阵进行修正,从而实现路径突变情况下的再次跟踪收敛。仿真实验结果验证了所提算法相比已有算法具有更高的正判率、更低的虚警率、更短的延迟帧数,可实现更好的再跟踪性能,从而提高路径突变场景下的声反馈抑制效果,并保证更好的音质。

1 AFC-PBFDFK 信号模型及路径突变情况分析

自适应滤波声反馈系统的信号流图如图1所示。传声器采集信号 $d(n)$ 由近端输入激励源信号 $s(n)$ 和反馈信号 $y(n)$ 构成。前向路径的放大增益为 K ,信号处理模块本身具有 Δt 时延,传声器信号经过前向路径后得到扬声器输出信号 $u(n)$ 。扬声器信号经过声学反馈路径 $h(n)$ 传递到传声器,得到反馈信号 $y(n)$ 。AFC系统通过自适应滤波器 $\hat{h}(n)$ 对声反馈路径 $h(n)$ 进行匹配估计,用于重构反馈信号估计,实现声反馈抑制。分区频域形式下,帧移量为 R ,快速傅里叶变换(Fast Fourier transform, FFT)长度为 M ,声学路径 $h(n)$ 被均分成长为 R 的 B 区分区滤波器,第 b

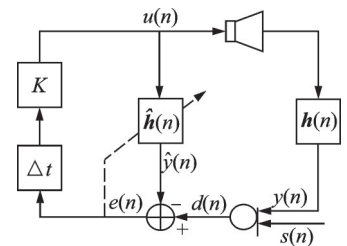


图1 AFC 信号流图

Fig.1 Diagram of AFC

区 ($b=0, 1, \dots, B-1$) 频域滤波器为 $H_b(k)$, 其中 k 表示经过分帧操作后的帧数。自适应滤波器 $\hat{h}(n)$ 也通过类似操作, 得到频域分块滤波器 $\hat{H}_b(k)$ 。第 b 区滤波器对应扬声器输入信号 $U_b(k)$, 近端输入激励源信号 $S(k)$, 传声器采集信号为 $D(k)$ 。状态转移方程和声反馈过程的观测方程分别为

$$H_b(k+1) = A[H_b(k) + \Delta H_b(k)] \quad (1)$$

$$D(k) = \sum_{b=0}^{B-1} G_{01, M-R} U_b(k) H_b(k) + S(k) \quad (2)$$

式中: A 为系统状态转移常数, 取值小于 1 且接近 1; $\Delta H_b(k)$ 为状态转移噪声; $G_{01, M-R} = F^{M \times M} \begin{bmatrix} O^{(M-R) \times (M-R)} & O^{(M-R) \times R} \\ O^{R \times (M-R)} & I^{R \times R} \end{bmatrix} (F^{M \times M})^{-1}$ 为消除圆周卷积混叠的约束矩阵; $I^{R \times R}$ 是大小为 $R \times R$ 的单位矩阵; $F^{M \times M}$ 为傅里叶变换矩阵。自适应滤波的先验误差信号频域表示为 $E(k)$ 。为简化计算, 将各矩阵均约束为对角矩阵, 由卡尔曼滤波理论得到 PBFDFKF 算法迭代公式。前向滤波以及滤波器参数更新过程为

$$E(k) = D(k) - \sum_{b=0}^{B-1} G_{01, M-R} U_b(k) \hat{H}_b(k) \quad (3)$$

$$K_b(k) = P_b(k) U_b^H(k) \left[U_b(k) P_b(k) U_b^H(k) + \frac{M}{R} \Phi_s(k) \right]^{-1} \quad (4)$$

$$\hat{H}_b(k+1) = A \left[\hat{H}_b(k) + G_{10, M-R} K_b(k) E(k) \right] \quad (5)$$

$$P_b(k+1) = A^2 \left[I^{M \times M} - \frac{R}{M} K_b(k) U_b(k) \right] P_b(k) + \Phi_\Delta(k) \quad (6)$$

式中: $\hat{H}_b(k)$ 为先验频域滤波器的估计; $\hat{H}_b(k+1)$ 为后验频域滤波器的估计; $E(k)$ 为先验误差信号, 上标 H 表示共轭转置操作; $P_b(k)$ 为状态误差协方差矩阵; $K_b(k)$ 为卡尔曼增益; $\Phi_s(k)$ 为观测噪声协方差矩阵, 对 $\Phi_s(k)$ 的估计可以通过平均估计误差信号的功率谱近似得到, 即 $\Phi_s(k) \approx \alpha \Phi_s(k-1) + (1-\alpha) |E(k)|^2$, 其中 α 为平滑因子, 取值范围是 $\alpha \in (0, 1)$; $\Phi_\Delta(k)$ 为状态转移噪声协方差矩阵, 对 $\Phi_\Delta(k)$ 的估计为

$$\Phi_\Delta(k) = (1 - A^2) \Phi_H(k) \approx (1 - A^2) \left| \hat{H}_b(k) \right|^2 \quad (7)$$

啸叫抑制系统对输入激励源信号的估计通过计算后验误差信号可得

$$E^+(k) = D(k) - \sum_{b=0}^{B-1} G_{01, M-R} U_b(k) \hat{H}_b(k+1) \quad (8)$$

对 AFC-PBFDFKF 算法的滤波器参数更新公式进行改写, 可以得到对应的频域变步长自适应滤波为

$$\hat{H}_b(k+1) = A \left[\hat{H}_b(k) + G_{10, M-R} \frac{\mu_b(k)}{U_b^H(k) U_b(k)} U_b^H(k) E(k) \right] \quad (9)$$

对应的变步长为 $\mu_b(k) = U_b(k) K_b(k)$, 相应状态误差协方差矩阵可写为

$$P_b(k+1) = A^2 \left(I^{M \times M} - \frac{R}{M} \mu_b(k) \right) P_b(k) + \Phi_\Delta(k) \quad (10)$$

以上矩阵 $P_b(k)$ 、 $\mu_b(k)$ 、 $\Phi_\Delta(k)$ 均为对角矩阵, 其对角元的第 i 个元素对应第 i 个频率点, 分别为 $P_{b,i}(k)$ 、 $\mu_{b,i}(k)$ 、 $\Phi_{\Delta,i}(k)$ 。根据文献[9]中的分析, 收敛稳定时, 对于 $A \rightarrow 1$ 的情况, 有 $P_{b,i}(\infty) \rightarrow 0$ 以及 $\mu_{b,i}(\infty) \rightarrow 0$, 即协方差误差矩阵对角元趋向于 0, 与之对应的变步长也趋向于 0, 这体现了收敛稳定时卡尔曼滤波的更新步长极小。若在稳态下发生路径突变, $A \rightarrow 1$ 则有 $\Phi_{\Delta,i}(\infty) \rightarrow 0$, 由式(10)的迭代公式

可以看出, $P_{b,i}(\infty) \rightarrow 0$ 几乎没有更新, 仍然维持在一个极小的数值, 导致对应的更新步长 $\mu_{b,i}(\infty)$ 也极小。但路径突变情况下需要较大的步长才能完成快速的跟踪与再次匹配, 这与当前卡尔曼滤波算法中的实际情况存在矛盾。这就是 AFC-PBFDKF 算法在路径突变情况下的锁死问题, 路径失配却无法通过自身迭代再次跟踪, 将会导致啸叫重现, 影响声反馈抑制算法的效果。

2 状态检测算法

2.1 加入状态检测模块的 AFC-PBFDKF 结构

卡尔曼滤波啸叫抑制状态检测算法基本结构如图 2 所示, 在得到先验误差信号后, 由卡尔曼滤波啸叫抑制模块得到的输入特征输入给卡尔曼滤波啸叫抑制状态检测算法 (Kalman-filter-based AFC with state detection model, KFSD), 最终输出值为 0 或 1 的状态检测二值变量 p_{flag} , 通过 p_{flag} 对卡尔曼滤波中的状态误差矩阵进行修正。若 p_{flag} 为 1, 则对状态估计误差协方差矩阵的赋值进行初始化, 反之则沿用之前由卡尔曼滤波计算得到的状态误差矩阵, 对应表达式为

$$P_b^+(k) = (1 - p_{\text{flag}}) \cdot P_b(k) + p_{\text{flag}} \cdot P_{\text{init}}(k) \quad (11)$$

式中 $P_{\text{init}}(k)$ 为初始化状态误差矩阵, 其对角元设置为一组较大的值, 用于修正卡尔曼滤波在路径突变情况下的步长。将经过状态检测模型修正后的状态估计误差协方差矩阵输入到卡尔曼滤波的迭代过程中, 得到相应的卡尔曼增益并进行滤波器的参数更新, 实现卡尔曼滤波啸叫抑制对路径突变的状态检测与再次跟踪收敛。

状态检测算法可以采取适当的传统信号处理策略, 例如设计某个与路径突变相关的计算量, 将其与预设阈值进行比较, 得到对应的状态检测结果。但阈值选取对算法效果有较大影响, 更低阈值可以使状态切换更加灵敏, 但同时会造成鲁棒性的下降。因此, 为实现更好的性能, 可以将上述的状态检测模块与 DNN 相结合。利用 DNN 的非线性建模能力以及在分类问题上的出色表现, 通过数据驱动的方式使状态检测效果得到提升。

2.2 轻量级的状态检测神经网络模块

采取一个轻量级神经网络模块用于进行路径突变的状态检测, 此前相关工作采取的网络输入特征为反馈损失增强向量 $r_{\text{de}}(k)$, 定义为采集信号与误差信号的窄带能量比的对数形式, 其第 i 个频点对应值的表达式为

$$r_{\text{de},i}(k) = \frac{\Phi_{D,i}(k)}{\Phi_{E,i}(k) + \epsilon} \quad (12)$$

式中 ϵ 为一个很小的正值, $\Phi_{D,i}(k)$ 和 $\Phi_{E,i}(k)$ 分别对应传声器采集信号能量 $\Phi_D(k)$ 与残差信号能量 $\Phi_E(k)$ 的第 i 个频点值, 表达式为

$$\Phi_D(k) = \alpha \Phi_D(k-1) + (1 - \alpha) |D(k)|^2 \quad (13)$$

$$\Phi_E(k) = \alpha \Phi_E(k-1) + (1 - \alpha) |E(k)|^2 \quad (14)$$

利用 $r_{\text{de}}(k)$ 在稳定状态下有较大值而在路径突变情况下会急剧下降的性质, 输入到网络中进行状态检测判断。命名上述算法为 KFSD-NN-v1, 对应的网络结构如图 3(a) 所示。输入特征为频域上分频带的 $L \times 1$ 维回声损失增强向量 $r_{\text{de}}(k)$, 其中 $L = M/2 + 1$ 是实序列 FFT 有效点数。前 4 个模块分别为

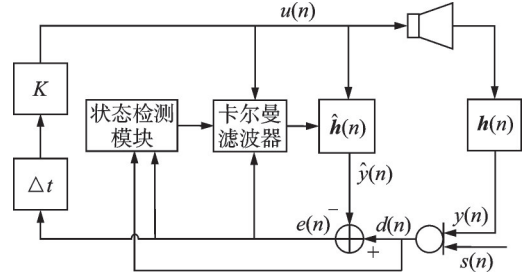


图 2 卡尔曼滤波啸叫抑制状态检测算法流程图
Fig.2 Diagram of Kalman-filter-based AFC with state detection model

$1 \times D$ 维度的全连接层(Full connection, FC)、 $D \times D$ 维度的门控循环单元(Gate recurrent unit, GRU)^[15] 以及 $D \times D$ 和 $D \times 1$ 维度的全连接层。对于分频带的 $L \times 1$ 维输入特征 $r_{de}(k)$, 前4个模块在频率维度 L 上共享相同的参数, 从而减小参数的存储量。全连接层之后的激活函数选取为参数整流线性单元(Parametric rectified linear unit, PReLU)^[16]。输入特征 $r_{de}(k)$ 通过 FC 层的升维和 PReLU 模块激活后, 进入 GRU 模块用于提取时序信息, 然后通过 FC 层降维和 PReLU 模块激活后得到 $L \times 1$ 维向量。第5个模块通过 $L \times 1$ 维 FC 模块将前面得到的分频带向量加权求和得到一个标量值, 最后通过 Sigmoid 函数模块, 得到一个 $(0, 1)$ 区间的网络输出值 $p_{\text{flag_out}}$ 。

为了充分利用卡尔曼滤波中各变量所隐含的信息, 网络输入特征为麦克风采集信号能量、残差信号能量和自适应滤波器更新能量这三者的并联, 定义为

$$\text{feat}(k) = \left[|E(k)|^2, |D(k)|^2, |\Delta\hat{H}_0(k)|^2, \dots, |\Delta\hat{H}_b(k)|^2, \dots, |\Delta\hat{H}_{B-1}(k)|^2 \right] \quad (15)$$

式中 $\Delta\hat{H}_b(k)$ 为第 b 区 ($b = 0, 1, \dots, B-1$) 频域自适应滤波器的更新量。这些变量通过神经网络的运算后, 输出路径突变状态检测的结果 $p_{\text{flag_out}}$ 。命名上述算法为 KFSD-NN-v2, 相比之前工作中基于规则定义的特征 $r_{de}(k)$, KFSD-NN-v2 可以利用包含更多信息的输入特征, 通过神经网络学习到更加丰富全面的内容, 提高检测的效果。KFSD-NN-v2 的网络结构如图 3(b) 所示, 整体结构与 KFSD-NN-v1 类似, 区别在于输入特征为分频带的 $L \times J$ 维向量 $\text{feat}(k)$, 其中 L 为频率维数, J 为并联后的通道维数。相应的, KFSD-NN-v2 中网络的第一层 FC 在通道维度上进行处理, 维度为 $J \times D$, 不同于 KFSD-NN-v1 中的第一层 FC 的维度 $1 \times D$ 。

以上两种网络结构的前4个模块均在频率维度 L 上共享相同的参数, 参数量得到了有效的减少, 但是较高的频率维度 L 可能会带来更高的运算量负担。因此, 进一步提出 KFSD-NN-v3, 在 KFSD-NN-

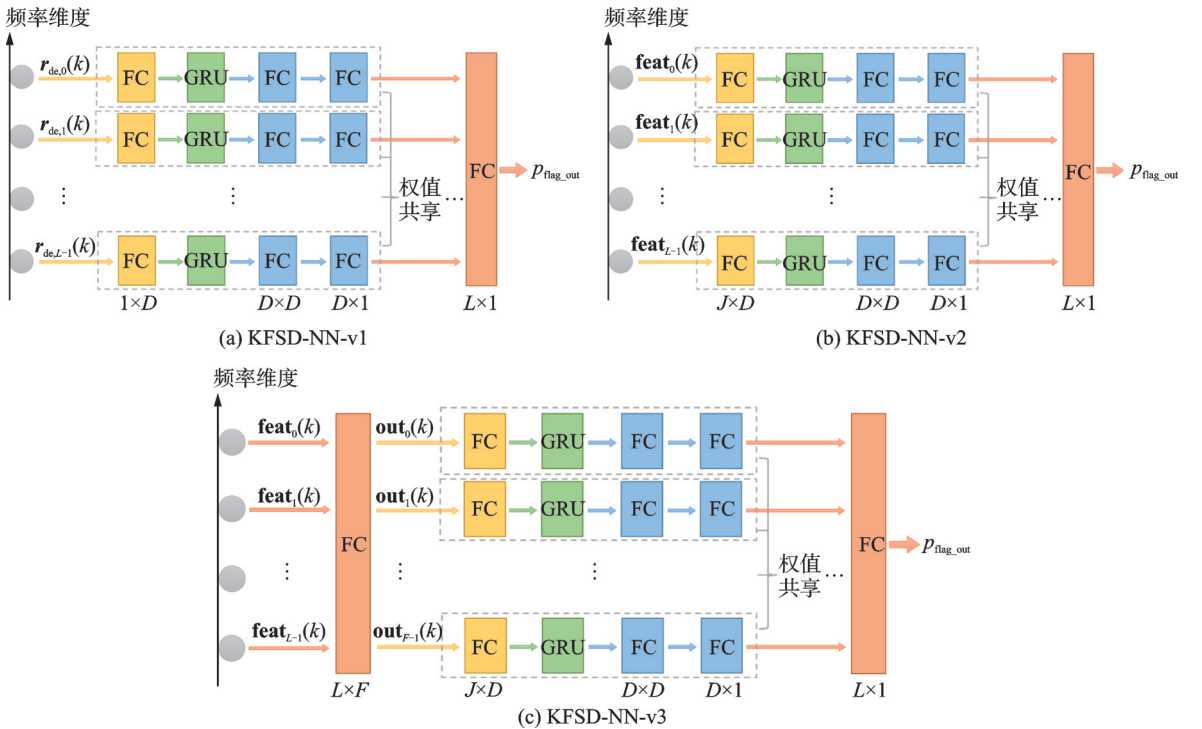


图3 KFSD-NN各算法的网络结构

Fig.3 Architectures of each algorithm for KFSD-NN

v2基础上对频率维度进行压缩并提取特征,网络结构如图3(c)所示。分频带的 $L \times J$ 维向量 $\mathbf{feat}(k)$ 经过 $L \times F$ 维的FC层后,得到经过频率维度压缩后的 $F \times J$ 维的中间输出特征 $\mathbf{out}(k)$,该FC层可以视为一个特征提取模块。然后将 $\mathbf{out}(k)$ 输入到与前面两种网络相类似的结构中,对应各模块的维度如图3(c)中所示。其中第2层FC模块在频率维度上共享,仅在通道维度上做处理,因此维度仍为 $J \times D$,代表通道维数从 J 变为 D ,用于后续操作。最后一层FC层在频率维度上操作,维度数为 $F \times 1$,不同于KFSD-NN-v2中的维度 $L \times 1$,同样也体现出KFSD-NN-v3对输入特征在频率维度上进行了压缩。与KFSD-NN-v1和KFSD-NN-v2相比较,KFSD-NN-v3通过压缩频率维度数,可以大幅度降低运算量。与此同时KFSD-NN-v3加入的频率维度压缩模块,即网络第一层维度为 $L \times F$ 的FC模块,会适当增加模型的参数量,但也带来了更强的频率压缩与特征提取的能力,可以更好地提取出有效信息用于后续的状态检测,从而实现更强的状态检测能力。

上述系统模型的训练流图如图4所示。为了模型的训练,需要构造路径突变状态检测的真实标签。设在训练的仿真过程中在第 k_c 帧将声反馈路径进行切换以模拟路径突变,则对应得到真实标签为

$$p_{\text{label}}(k) = \begin{cases} 1 & k_c + k_1 \leq k < k_c + k_1 + k_s \\ 0 & \text{其他} \end{cases} \quad (16)$$

式中: k_1 为状态检测模型的检测延迟帧数; k_s 为标签为1的部分持续的帧数。在仿真过程中,将已知的近端信号输入给融合神经网络的卡尔曼滤波啸叫抑制状态检测模型,利用从16 kHz仿真房间常数数据集SLR26^[17]中选取的声反馈路径RIR,仿真构造出卡尔曼声反馈抑制的流程。在时间维度上随机选取 m 帧,在该帧的起始将声反馈路径从一条切换到另一条以模拟路径突变。对应得到真实标签 $p_{\text{label}}(k)$,在其监督下对卡尔曼滤波啸叫抑制模块进行修正。然后将卡尔曼滤波啸叫抑制模块得到的变量经过计算,得到状态检测模型的输入特征,在真实标签的监督下将其输入到神经网络,神经网络输出得到 $p_{\text{flag_out}}$ 。训练目标是使网络输出接近真实标签,对应代价函数选取为二值交叉熵(Binary cross entropy, BCE),即

$$\text{Loss} = \text{BCE}(p_{\text{flag_out}}, p_{\text{label}}) \quad (17)$$

通过数据驱动式的训练,可以使网络输出 $p_{\text{flag_out}}$ 尽量逼近目标的真实标签 p_{label} ,从而实现状态检测。

在推断环节,将得到的预训练模型赋值给状态检测模型。推断过程中,由卡尔曼滤波啸叫抑制模块得到的输入特征输入给神经网络,得到在时间序列上某一时刻的神经网络输出为 $p_{\text{flag_out}}$,转化为状态检测的二值变量 p_{flag} ,即

$$p_{\text{flag}} = \begin{cases} 1 & p_{\text{flag_out}} \geq 0.5 \\ 0 & \text{其他} \end{cases} \quad (18)$$

利用 p_{flag} ,可按照式(11)对卡尔曼滤波中的状态误差矩阵进行修正。

3 仿真实验与结果讨论

3.1 实验设置

啸叫抑制方法采取分区频域卡尔曼滤波。采样率为16 kHz, $M = 512$,帧移量为 $R = 256$,声反馈路径房间冲激响应RIR的阶数为 $N = 1024$,对应分区数 $B = 4$ 。前向时延 $\Delta t = 512$ 对应延时为32 ms,前

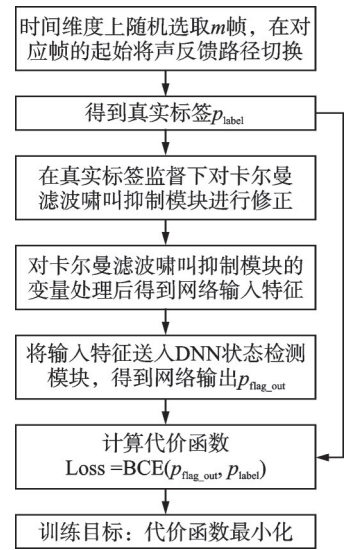


图4 训练流程图
Fig.4 Flowchart of training stage

向增益为 $K=6$ 。卡尔曼滤波中转移系数 $A=0.999\ 99$,各功率谱估计的平滑因子 $\alpha=0.7$,状态误差协方差矩阵对角元各分区初始值分别为 $0.8,0.4,0.2$ 和 0.1 。神经网络输入特征的频率维度 $L=M/2+1=257$,KFSD-NN-v2 和 KFSD-NN-v3 输入特征的分块维度 $J=6$,KFSD-NN-v3 在频率维度压缩后的维度 $F=32$,FC 层维度 $D=8$,GRU 层数为 1 ,隐藏单元数 $D=8$ 。

训练数据和测试样本均来源于 AEC-Challenge 合成数据集^[18],选取远端信号作为输入的激励源信号,其中训练集数据的样本量为 $5\ 000$,验证集数据的样本量为 300 ,测试集数据的样本量为 500 。声反馈路径 RIR 从 SLR26 数据集中选取^[17],选用 80 个房间中的 RIR 用于训练和验证,另外 20 个房间中的 RIR 用于测试,均截取到 $1\ 024$ 阶。在时间维度上随机选取 $0\sim 3$ 帧(0 帧代表 RIR 没有发生变化的情况),在对应帧将 RIR 从一条随机切换到另一条,模拟路径突变。真实标签设置中, $k_t=3,k_s=5$ 。

本文选取的评价指标为语音质量感知评估(Perceptual evaluation of speech quality,PESQ)^[19]和尺度不变信号失真比(Scale-invariant signal-to-distortion ratios,SISDR)^[20]。定义正判率(True positive rate,TPR)为状态检测判断结果中正确检测出的状态突变个数占总状态突变真实标签个数的比例;虚警率(False alarm rate,FAR)为状态检测判断结果中错误判断的状态检测个数占总状态检测判断个数的比例。延迟时间(Latency)为状态检测判断帧与真实标签帧相比的平均延迟(单位为帧)。归一化适配率(Normalized misalignment,NMIS)为估计路径与实际路径在每一点上的差距,单位为 dB,定义为

$$\text{NMIS} = 10 \lg \left(\frac{\sum_{i=0}^N (h_i - \hat{h}_i)^2}{\sum_{i=0}^N (h_i)^2} \right) \quad (19)$$

3.2 实验结果

将本文提出的融合神经网络的状态检测模型 KFSD-NN-v2 和 KFSD-NN-v3 与此前提出的 KFSD-NN-v1、文献[10]中基于阴影滤波器的状态检测模型 KFSD-shadow 以及不加入状态检测模型的卡尔曼滤波结构 KFSD-none 进行对比,在 500 条测试样本上进行评估并取平均,结果如表 1 所示。从表 1 可以看出,所提算法在 SISDR、PESQ、TPR、FAR 以及 Latency 各指标上都优于 KFSD-shadow。与 KFSD-NN-v1 相比,本文提出的 KFSD-NN-v2 和 KFSD-NN-v3 在 SISDR 和 PESQ 指标上也有所提升。KFSD-NN-v2 轻微牺牲了检测的延迟速度,提高了检测的正判率,降低了误检率,从而提高了音质,得到了最优的 PESQ 和次优的 SISDR,证明了 KFSD-NN-v2 具有良好的特征提取能力和稳定的检测性能。KFSD-NN-v3 在检测延迟速度上取得了最优结果,平均仅需 1.33 帧即可检测出路径突变,从而得到最优的 SISDR 和次优的 PESQ 结果,但在正判率和误检率上的表现稍微逊色。模型的运算量由表 1 中的 FLOP(Floating point operation)来表示,可以看出,KFSD-NN-v1 和 KFSD-NN-v2 具有较小的参数量,KFSD-NN-v3 增加了参数量但减少了运算量,KFSD-NN-v3 的运算量与传统算法 KFSD-shadow 的运算量更为接近,相比之下增加幅度较小。总体来看,本文提出的神经网络状态检测模型在参数量和运算量上均为轻量级,在实时系统上部署具有可行性。

表 1 各模型的客观评价指标测试结果

Table 1 Objective evaluation index test results for all model

模型	SISDR	PESQ	TPR/%	FAR/%	Latency/帧	参数量	FLOP/ 10^6
KFSD-NN-v1	2.55	2.53	<u>98.20</u>	<u>9.24</u>	<u>2.95</u>	790	19.31
KFSD-NN-v2	3.33	2.57	99.80	2.92	3.86	<u>830</u>	20.59
KFSD-NN-v3	3.87	<u>2.56</u>	97.40	9.81	1.33	8 860	<u>8.78</u>
KFSD-shadow	-7.90	2.14	92.40	53.56	7.67	—	5.83
KFSD-none	-30.54	1.28	—	—	—	—	—

注:黑体数字表示最优,下划线数字表示次优。

图5给出了基于KFSD-NN的3种模型与KFSD-shadow以及KFSD-none在一条测试样本上的语谱图结果。声反馈的路径突变发生在第4s,用于测试的两条房间冲激响应RIR是在混响时间约为0.5s的8m×6m×3m房间里通过移动麦克风相对扬声器位置实际采集得到,均截取到1024阶。从图5中可以看出,不加入状态检测模型的KFSD-none在路径突变后发生了严重的啸叫现象;加入阴影滤波器的KFSD-shadow算法可以进行状态检测,但是由于检测延迟较高,仍有明显的啸叫部分持续了超过0.5s;而所提出3种模型相比KFSD-shadow具有更快的状态检测速度,啸叫部分相比KFSD-shadow更不明显,并且啸叫部分的持续时间在KFSD-NN-v1、KFSD-NN-v2、KFSD-NN-v3模型的结果中递减,证明了所提算法具有更加快速实现再次跟踪收敛与啸叫抑制的优势。图6给出了各模型对应的NMIS随时间变化的曲线图,其结果与语谱图的表现一致,也展示出所提算法具有更快的检测速度,从而得到更优的状态检测及啸叫抑制效果。

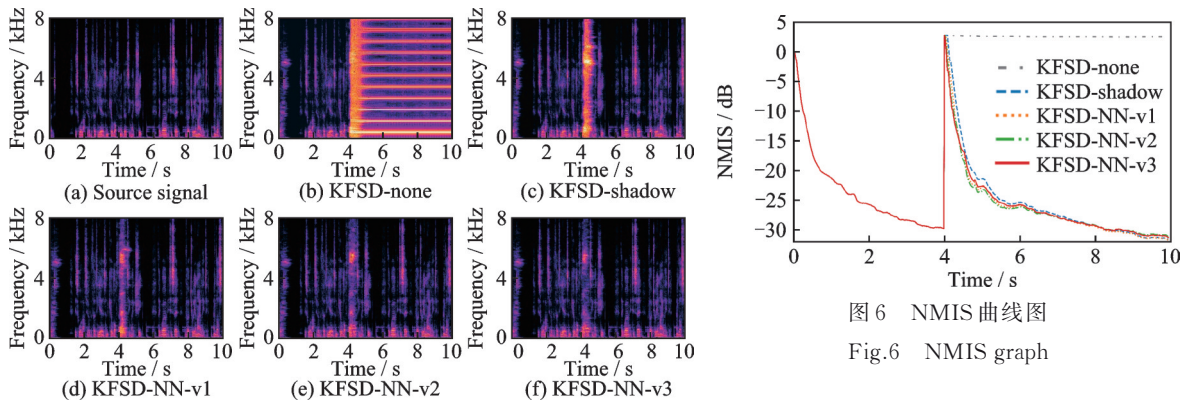


图5 语谱图结果

Fig.5 Spectrogram results

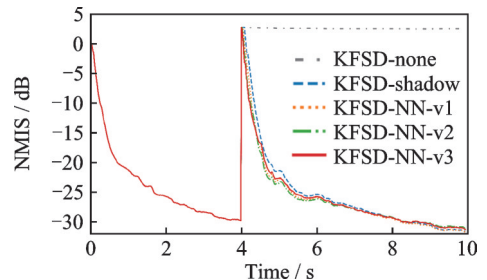


图6 NMIS 曲线图

Fig.6 NMIS graph

4 结束语

本文针对声反馈抑制卡尔曼滤波在路径突变时无法再次跟踪收敛的问题,提出了一种融合轻量级神经网络的状态检测模型,进而提出了KFSD-NN系列算法。所提算法结合深度神经网络和数据驱动的优势,通过使用规模较小且计算量较低的深度神经网络,将卡尔曼滤波声反馈抑制系统的传声器采集信号、残差信号和滤波器更新量作为输入特征,通过神经网络对卡尔曼滤波的状态误差协方差矩阵进行修正,实现路径突变情况下的再次跟踪和收敛。经过在数据集上的训练与测试,并和以往的阴影滤波器状态切换算法进行比对,验证了所提算法的优势。仿真实验结果表明,所提算法具有较高的正判率、较低的虚警率和较短的延迟帧数,可以在路径突变发生后迅速做出状态检测,进入再跟踪状态,迅速抑制路径突变导致的啸叫,保证了音质的稳定。算法利用较小规模和较低计算量的深度神经网络,得到了更加准确和迅速的状态检测,实现了良好的快速再跟踪性能。

参考文献:

- [1] RICHARD D. Fifty years of acoustic feedback control: State of the art and future challenges[J]. Proceedings of the IEEE, 2011, 99(2): 288-327.
- [2] SIQUEIRA M G, ALWAN A. Steady-state analysis of continuous adaptation in acoustic feedback reduction systems for hearing-aids[J]. IEEE Transactions on Speech and Audio Processing, 2000, 8(4): 443-453.
- [3] ROMBOUTS G, VAN W T, STRUYVE K, et al. Acoustic feedback cancellation for long acoustic paths using a nonstationary source model[J]. IEEE Transactions on Signal Processing, 2006, 54(9): 3426-3434.
- [4] SPRIET A, DOCLO S, MOONEN M, et al. Springer handbook of speech processing[M]. Berlin, Heidelberg: Springer, 2008: 979-1000.

- [5] ENZNER G, VARY P. Frequency-domain adaptive Kalman filter for acoustic echo control in hands-free telephones[J]. *Signal Processing*, 2006, 86(6): 1140-1156.
- [6] MALIK S, ENZNER G. State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation[J]. *IEEE Transactions on Audio, Speech, and Language Processing*, 2012, 20(7): 2065-2079.
- [7] BERNARDI G, WATERSCHOOT T V, WOUTERS J, et al. A PEM-based frequency-domain Kalman filter for adaptive feedback cancellation[C]//*Proceedings of the 2015 23rd European Signal Processing Conference (EUSIPCO)*. Nice, France: IEEE, 2015: 270-274.
- [8] BERNARDI G, VAN W T, WOUTERS J, et al. Adaptive feedback cancellation using a partitioned-block frequency-domain Kalman filter approach with PEM-based signal prewhitening[J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2017, 25(9): 1784-1798.
- [9] YANG F, ENZNER G, YANG J. Frequency-domain adaptive Kalman filter with fast recovery of abrupt echo-path changes[J]. *IEEE Signal Processing Letters*, 2017, 24(12): 1778-1782.
- [10] LU C, YANG F, YANG J. An adaptive time-domain Kalman filtering approach to acoustic feedback cancellation for hearing aids[J]. *Chinese Journal of Electronics*, 2020, 29(1): 139-146.
- [11] TONG S, GU H, YU K. A comparative study of robustness of deep learning approaches for VAD[C]//*Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai, China: IEEE, 2016: 5695-5699.
- [12] XIA X, TOGNERI R, SOHEL F, et al. A survey: Neural network-based deep learning for acoustic event detection[J]. *Circuits, Systems, and Signal Processing*, 2019, 38(8): 3433-3453.
- [13] CHEN Z, HAO Y, CHEN Y, et al. A neural network-based howling detection method for real-time communication applications[C]//*Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore: IEEE, 2022: 206-210.
- [14] GUO H, LE X, CHEN K, et al. A light-weight state detection model for Kalman filter based acoustic feedback cancellation with rapid recovery from abrupt path changes[C]//*Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Seoul, Korea: IEEE, 2024: 456-460.
- [15] CHO K, MERRIENBOER B V, GULCEHRE Ç, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[C]//*Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Doha, Qatar: ACL, 2014: 1724-1734.
- [16] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//*Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile: IEEE, 2015: 1026-1034.
- [17] KO T, PEDDINTI V, POVEY D, et al. A study on data augmentation of reverberant speech for robust speech recognition [C]//*Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. New Orleans, LA, USA: IEEE, 2017: 5220-5224.
- [18] CUTLER R, SAABAS A, PARNAMAA T, et al. ICASSP 2022 acoustic echo cancellation challenge[C]//*Proceedings of the 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Singapore: IEEE, 2022: 9107-9111.
- [19] RIX A W, BEERENDS J G, HOLLIER M P, et al. Perceptual evaluation of speech quality (PESQ)—A new method for speech quality assessment of telephone networks and codecs[C]//*Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings*. Salt Lake City, UT, USA: IEEE, 2001: 749-752.
- [20] ROUX J L, WISDOM S, ERDOGAN H, et al. SDR-half-baked or well done? [C]//*Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, UK: IEEE, 2019: 626-630.

作者简介:



郭昊诚(2000-),男,硕士研究生,研究方向:音频信号处理、啸叫抑制,E-mail: guohaocheng@smail.nju.edu.cn。



陈楷(1980-),通信作者,男,博士,副教授,研究方向:声场调控、声信号增强,E-mail: chenka@nju.edu.cn。



卢晶(1977-),通信作者,男,博士,教授,研究方向:声场调控、声信号增强,E-mail: lujing@nju.edu.cn。

(编辑:王婕,张黄群)