Table 1: The results of two different transfer modes.

Transferred layer	***Fixed***		***Fine-tuning***	
	F-score	*N.params*	*F-score*	*N.params*
L_1	91.9%	20.58K	**93.2%**	20.72K
L_2	91.7%	13.33K	92.0%	20.72K
L_3	91.1%	13.33K	91.7%	20.72K
L_{all}	82.6%	5.79K	92.3%	20.72K

Table 2: The results of two different transfer strategies.

Transfer	*Fixed*			*Fine-tuning*		
layer No.	*P*	*R*	*F-score*	*P*	*R*	*F-score*
Entire	79.3%	86.2%	82.6%	89.0%	95.7%	92.3%
L_1	88.9%	95.1%	91.9%	**90.1%**	**96.0%**	**93.2%**
L_2	88.2%	95.5%	91.7%	88.8%	95.5%	92.0%
L_3	87.4%	95.1%	91.1%	88.5%	95.2%	91.7%

Table 3: The detection results in frame-level.

Polyphonic	*Frames*		*Baseline*			*Transfer learning*		
song	*off*	*on*	*P* (%)	*R* (%)	*F* (%)	*P* (%)	*R* (%)	*F* (%)
No.1	3390	6098	79.6	91.4	85.1	84.1	91.8	**87.8**
No.2	5844	8366	96.4	93.2	89.7	88.4	92.9	**90.6**
No.3	2744	4793	84.5	92.3	88.2	86.5	91.7	**89.1**
No.4	6423	2911	89.5	94.4	**91.9**	86.7	93.7	90.1
No.5	1475	4561	90.2	94.3	92.2	91.0	97.8	**94.2**
No.6	4945	5754	86.5	91.6	89.0	89.3	96.5	**92.8**
No.7	3218	7220	96.5	95.7	96.1	95.8	97.7	**96.8**
No.8	2458	9922	66.6	89.9	76.5	70.8	91.4	**79.8**
No.9	2938	5384	82.8	85.4	84.1	92.9	97.3	**95.1**
No.10	4166	7476	83.0	90.0	86.3	89.6	98.6	**93.9**

Table 4: The detection results in frame-level.

Polyphonic	***Frames***		***Baseline***			***Transfer learning***		
song	*off*	*on*	*P* (%)	*R* (%)	*F* (%)	*P* (%)	*R* (%)	*F* (%)
No.1	2938	5384	82.8	85.4	84.1	92.9	97.3	**95.1**
No.2	4166	7476	83.0	90.0	86.3	89.6	98.6	**93.9**
•••	•••	•••	•••	•••	•••	•••	•••	•••
No.60	3218	7220	96.5	95.7	96.1	95.8	97.7	**96.8**
Overall			86.1	93.2	89.5	90.1	96.0	**93.2**