	Subjective preference (%) in naturalness					
Language	**LSTM**	**Concat**	**WaveNet** (L)	**WaveNet** (L+F)	No preference	p value
North	23.3	**63.6**			13.1	$\ll 10^{-9}$
American	18.7		**69.3**		12.0	$\ll 10^{-9}$
English	7.6			**82.0**	10.4	$\ll 10^{-9}$
		32.4	**41.2**		26.4	0.003
		20.1		**49.3**	30.6	$\ll 10^{-9}$
			17.8	**37.9**	44.3	$\ll 10^{-9}$
Mandarin	**50.6**	15.6			33.8	$\ll 10^{-9}$
Chinese	25.0		23.3		51.8	0.476
	12.5			**29.3**	58.2	$\ll 10^{-9}$
		17.6	**43.1**		39.3	$\ll 10^{-9}$
		7.6		**55.9**	36.5	$\ll 10^{-9}$
			10.0	**25.5**	64.5	$\ll 10^{-9}$

Table 2: Subjective preference scores of speech samples between LSTM-RNN-based statistical parametric (**LSTM**), HMM-driven unit selection concatenative (**Concat**), and proposed WaveNet-based speech synthesizers. Each row of the table denotes scores of a paired comparison test between two synthesizers. Scores of the synthesizers which were significantly better than their competing ones at $p < 0.01$ level were shown in the bold type. Note that **WaveNet** (L) and **WaveNet** (L+F) correspond to WaveNet conditioned on linguistic features only and that conditioned on both linguistic features and F_0 values.