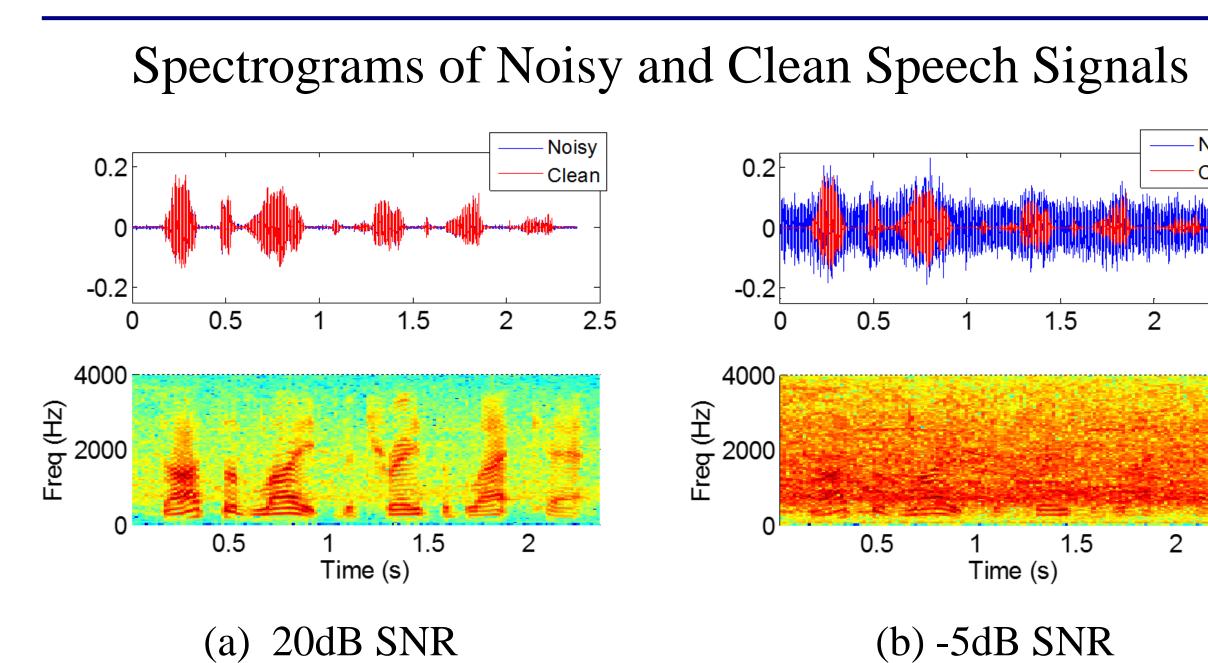


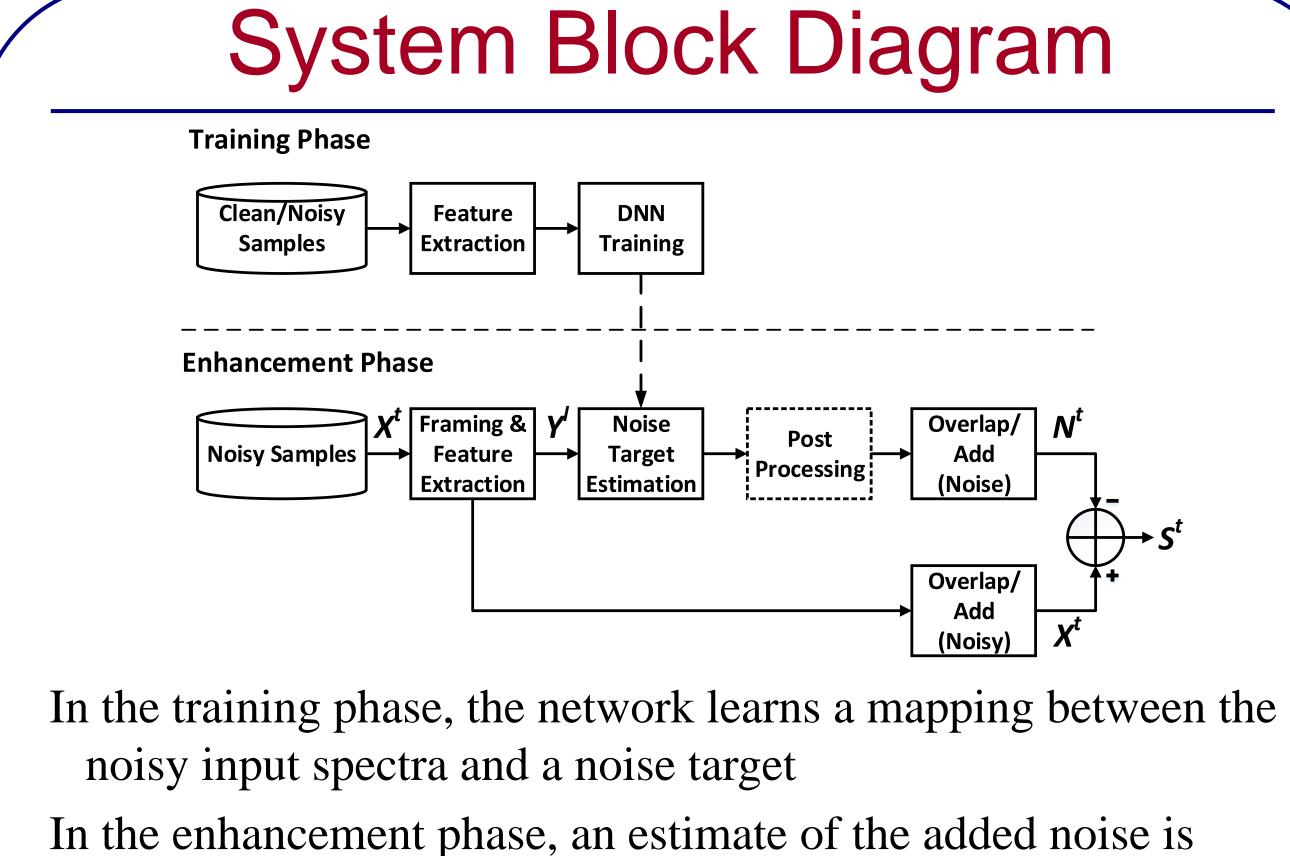
## Objective

- Evaluate the performance of different training targets for deep neural network (DNN) speech enhancement based on noise prediction
- Compare the performance of the speech enhancement systems based on noise prediction to that of a conventional SE system based on prediction of clean speech.

### **Noise Prediction Rationale**



- The speech signal is dominated by the noise at average SNR lower than 0dB
- > Learn a mapping between the noisy signal input and the added noise [1]
- The noisy signal phase is dominated by the phase of the noise at low SNR values
  - $\blacktriangleright$  Use the noisy signal phase to reconstruct the noise



reconstructed, and the noise-free speech is obtained by time domain subtraction

# A Study of Training Targets for Deep Neural Network-Based Speech Enhancement Using Noise Prediction

- 1. Log  $\succ$  S
- 2. For

WI 3. No:

W an  $\succ$  Eq

Noise-fre Corpus Training types Performa

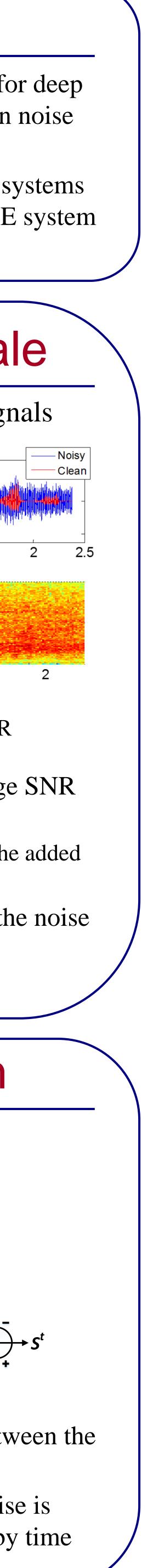
conver model

<b>Training Targets</b>	Results									
og Magnitude Spectrum (LogFFT)			SNR	ът •		T DDD				
STFT magnitude spectrum is log compressed			(dB)	Noisy	NAT	LogFFT	FFT-MASK	NRM		
ourier Magnitude Spectrum Mask (FFT-MAS	SK)		20	3.182	3.426	3.292	3.413	3.530		
			15	2.875	3.242	3.007	3.134	3.266		
Ratio of the noise and noisy speech magnitude spe	ectra		10	2.569	3.020	$\frac{2.715}{2.420}$	$\frac{2.842}{2.528}$	2.976		
$M_{FFT}(t,\omega) = \frac{N(t,\omega)}{X(t,\omega)}$			5	$\begin{array}{c c} 2.288 \\ \hline 2.036 \end{array}$	$\begin{array}{r} 2.760 \\ 2.475 \end{array}$	$\begin{array}{r} 2.420 \\ \hline 2.137 \end{array}$	$\begin{array}{r} 2.538 \\ \hline 2.233 \end{array}$	$2.664 \\ 2.345$		
$XIFF(t,\omega) = \overline{X(t,\omega)}$			-5	$\frac{2.030}{1.779}$	2.475	$\frac{2.137}{1.865}$	$\frac{2.233}{1.942}$	2.032		
$\mathbf{V}$			AVG.	2.455	2.851	$\frac{1.000}{2.573}$	2.684	2.802		
where $N(t,\omega)$ and $X(t,\omega)$ are respectively the respectra of the added noise and noisy speech	nagmtude	PESQ	) scores	for pro	posed a	and NAT	systems in ur	seen nois		
Joise Ratio Mask (NRM)			SNR							
$NRM(t,\omega) = \left(\frac{N^2(t,\omega)}{S^2(t,\omega) + N^2(t,\omega)}\right)^{\frac{1}{2}}$			(dB) 20	Noisy           0.958	NAT 0.935	LogFFT 0.965	FFT-MASK 0.965	NRM 0.970		
			15	0.925	0.922	0.937	0.939	0.949		
where $N^2(t,\omega)$ and $S^2(t,\omega)$ are respectively the	e added noise		10	0.876	0.900	0.893	0.899	0.915		
and speech power spectral densities.			5	0.813	0.862	0.832	0.842	0.863		
Equivalent to a frequency domain square root Wier	ner filter		0	0.736	0.804	0.755	0.767	0.793		
			-5	0.650		0.666	0.677	0.706		
Experiments		STOI	AVG.	0.826 for proj	0.858 posed a	0.841 .nd NAT s	0.848 Systems in un	0.866 seen noise		
free speech and noise data were obtained from ous and non-speech sound database respective		Observations								
ng datasets of about 50 hours were made by a	dding 50 noise	In seen n	oise:							
s to clean speech; testing was done with 10 nd	oise types	• The n	oise pre	diction	models	in gener	al, perform v	vell in		
mance of noise prediction models compared t			<b></b>			of noisy s	<b>A</b>			
			U		•	•	•	1 1 •		
entional, noise-aware trained (NAT) [2], spee el	ech prediction		RM monoclassical RM mon	<b>A</b>		best amon	g the noise m	nodels 1n		
		• The N	JRM out	tperform	ns the N	NAT mode	el on average	; however		
Results				e latter	at low S	SNR valu	es			
CND		In unseei								
$\left \begin{array}{c c} SNR \\ (dB) \end{array}\right  Noisy \left \begin{array}{c c} NAT \end{array}\right  LogFFT \left \begin{array}{c} FFT-MASK \end{array}\right $	NRM	• The N	VRM is t	the best	among	the noise	models in er	hancing		
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	3.765	the qu	ality an	d intelli	gibility	of noisy	speech			
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	3.590	• The NRM model performs better than the NAT model in								
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	3.380	enhancing quality at higher SNR values but is worse at low								
5  2.072  3.114  2.982  2.975	3.134		values a	<b>C</b>	•					
0 1.791 <b>2.932</b> 2.708 2.665	2.845				U					
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	2.513			<b>–</b>			the NAT m			
AVG.     2.246     3.153     3.084     3.076	3.205	enhan	icing int	elligibil	ity at h	igher SNI	R values and	on averag		
SQ scores for proposed and NAT systems in s	seen noise	is slig	shtly wo	rse at lo	wer SN	VR values				
SNR (dB)NoisyNATLogFFTFFT-MASK	NRM		Summary							
20 0.961 0.937 <b>0.981</b> 0.974	0.977						-			
15 0.926 0.928 <b>0.968</b> 0.958	0.962	The NRM	A was th	ne best a	all-roun	d noise ta	rget. It outpe	erformed t		
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	0.941						d in improvi			
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	0.910						hort at lower	U		
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	0.868	memg	JUIIILY II	1 1115551	1 110180,		non al IUWEI	SINIX Väl		
-5         0.608         0.834         0.817         0.787           AVG.         0.812         0.897         0.917         0.901	$ \begin{array}{c c} 0.808 \\ \hline 0.911 \end{array} $	REFERENC	ES							
				.V. Andersc	on, IEEE In	t'l Conf. on M	Iachine Learning ar	d Application		
M scores for proposed and NAT systems in s	een noise				1 1	T D				

**PES** 

Training Targets						Results							
g Magn <sup>2</sup>	itude Spe	ctrum (LogF	FT)				SNR	Naiger					
TFT ma	gnitude sp	ectrum is log c	compressed				(dB)	Noisy	NAT	LogFFT	FFT-MASK	NRM	
rier M	aonitude	Spectrum Ma	ask (FFT-MAS	SK)			20	3.182	3.426	3.292	3.413	3.530	
	•	-					15	2.875	3.242	$\frac{3.007}{0.715}$	3.134	3.266	
	ne noise ai	nd noisy speeci	h magnitude spe	ectra			10	$\begin{array}{r} 2.569 \\ \hline 2.288 \end{array}$	$\begin{array}{c} 3.020 \\ \hline 2.760 \end{array}$	$\begin{array}{r} 2.715 \\ \hline 2.420 \end{array}$	$\begin{array}{c c} 2.842 \\ \hline 2.538 \end{array}$	2.976 2.664	
		$M_{FFT}(t,\omega) = \frac{N}{N}$	$\overline{\mathrm{V}(t,\omega)}$				$\begin{array}{c c} 5\\ \hline 0 \end{array}$	2.288 2.036	$\frac{2.700}{2.475}$	$\frac{2.420}{2.137}$	$\frac{2.338}{2.233}$	2.004 2.345	
		Y = X	$X(t,\omega)$		- i i		-5	1.779	2.182	$\frac{2.191}{1.865}$	$\frac{2.233}{1.942}$	2.032	
no N(1)	ta) and I	$I(t \alpha)$ are real	postivaly that	magnituda			AVG.	2.455	2.851	2.573	2.684	2.802	
a of	the adde	d noise and n	pectively the r noisy speech	magintuue		PESÇ	) scores	for pro	posed a	and NAT	systems in ur	nseen noi	S
Rati	o Mask (	NRM)					SNR						
		$\sim$ ( $N^2$	$^{2}(t,\omega)$ $\rangle^{\frac{1}{2}}$				(dB)	Noisy	NAT	LogFFT	FFT-MASK	NRM	
$NRM(t,\omega) = \left(\frac{N^2(t,\omega)}{S^2(t,\omega) + N^2(t,\omega)}\right)^{\frac{1}{2}}$				20	0.958	0.935	0.965	0.965	0.970				
$\mathbf{V}$ have $\mathbf{V}^2(1, \mathbf{x})$ and $\mathbf{C}^2(1, \mathbf{x})$ are noticed to the odded matrix				15	0.925	0.922	0.937	0.939	0.949				
here $N^2(t,\omega)$ and $S^2(t,\omega)$ are respectively the added noise				10	0.876	0.900	0.893	0.899	0.915				
ec	h power s	spectral densi	ities.					0.813	0.862	0.832	0.842	0.863	
ent	t to a frequ	ency domain s	quare root Wier	ner filter				0.736	$\begin{array}{ c c c c }\hline 0.804 \\\hline 0.727 \\\hline \end{array}$	0.755	$\begin{array}{c c} 0.767 \\ \hline 0.677 \end{array}$	0.793	
_		-					-5 AVG.	$ \begin{array}{c c} 0.650 \\ \hline 0.826 \end{array} $	<b>0.727</b> 0.858	0.666 0.841	$\begin{array}{c c} 0.677 \\ \hline 0.848 \end{array}$	0.706 0.866	
	E	xperim	ents			STOI					systems in un		3(
d no	on-speech	sound databa	e obtained from ase respective ere made by a	ly	se	In seen n	oise:	0	bse	rvatio	DNS		-
	-	-	one with 10 nc els compared te	• -			<b></b>			, in gener of noisy s	al, perform v speech	vell in	
lonal, i	noise-aw	are trained (N	NAT) [2], spee	ch prediction			VRM monocology			est amon	g the noise m	nodels in	
		Resul	ts					•		NAT mode SNR valu	el on average es	; howeve	۱` ۲.
		110001			-	In unsee	n noise:						
									amona	the noise	models in er	nhancing	-
NR lB)	•	IAT LogFFT		NRM					•			mancing	
3) )	3.027 3.	.506 3.686	3.720	3.765		the qu	ality an	d intell	igibility	of noisy	speech		
	3.027     3.       2.701     3.	.506 3.686 .394 3.481	3.720 3.511	3.765 3.590		the que The N	ality an	d intell	igibility forms b	of noisy better than	speech the NAT mo	odel in	
<b>B</b> )	3.027     3.       2.701     3.       2.380     3.	$\begin{array}{c c}     506 & 3.686 \\     394 & 3.481 \\     265 & 3.240 \\   \end{array}$	$ \begin{array}{c c} \hline 3.720 \\ \hline 3.511 \\ \hline 3.258 \\ \hline \end{array} $	3.765 3.590 3.380		the que The N	ality an	d intell	igibility forms b	of noisy better than	speech	odel in	e
	3.027       3.         2.701       3.         2.380       3.         2.072       3.	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c c} \hline 3.720 \\ \hline 3.511 \\ \hline 3.258 \\ \hline 2.975 \\ \hline \end{array} $	3.765 3.590 3.380 3.134		the que The N enhan	ality an	d intelle odel per ality at	igibility forms b higher \$	of noisy better than	speech the NAT mo	odel in	e
)	3.027       3.         2.701       3.         2.380       3.         2.072       3.         1.791       2.	.506       3.686         .394       3.481         .265       3.240         .114       2.982 <b>932</b> 2.708	$\begin{array}{c c} & 3.720 \\ \hline 3.511 \\ \hline 3.258 \\ \hline 2.975 \\ \hline 2.665 \end{array}$	3.765         3.590         3.380         3.134         2.845		the que The N enhan SNR	ality an JRM mo icing qu values a	d intelled odel per ality at and on a	igibility forms b higher s verage	of noisy better than SNR valu	speech the NAT mo es but is wor	odel in se at low	e
	3.027       3.         2.701       3.         2.380       3.         2.072       3.         1.791 <b>2.</b> 1.503 <b>2.</b>	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	$ \begin{array}{c c} \hline 3.720 \\ \hline 3.511 \\ \hline 3.258 \\ \hline 2.975 \\ \hline \end{array} $	3.765 3.590 3.380 3.134		the que The N enhan SNR The N	ality an NRM mo icing qu values a NRM mo	d intelle odel per ality at nd on a odel per	igibility forms b higher s verage forms b	of noisy better than SNR valu	speech the NAT mo es but is wor	odel in se at low odel in	
•	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	.506       3.686         .394       3.481         .265       3.240         .114       2.982 <b>932</b> 2.708 <b>708</b> 2.409         .153       3.084	$\begin{array}{c c} & 3.720 \\ \hline 3.511 \\ \hline 3.258 \\ \hline 2.975 \\ \hline 2.665 \\ \hline 2.327 \end{array}$	3.765         3.590         3.380         3.134         2.845         2.513         3.205		the quantum the quantum the quantum the quantum term of the N enhant SNR for the N enhant term of term	ality an NRM mo icing qu values a NRM mo icing int	d intelle odel per ality at nd on a odel per elligibi	igibility forms b higher s verage forms b lity at h	of noisy better than SNR valu	speech the NAT mo es but is wor the NAT mo R values and	odel in se at low odel in	
B) 20 5 0 5 0 5 7 G. VG.	$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	.506       3.686         .394       3.481         .265       3.240         .114       2.982 <b>932</b> 2.708 <b>708</b> 2.409         .153       3.084	3.720 3.511 3.258 2.975 2.665 2.327 3.076 <b>X systems in s</b>	3.765         3.590         3.380         3.134         2.845         2.513         3.205		the quantum the quantum the quantum the quantum term of the N enhant SNR for the N enhant term of term	ality an NRM mo icing qu values a NRM mo icing int	d intelle odel per ality at nd on a odel per elligibit rse at lo	igibility forms b higher S verage forms b lity at h ower SN	of noisy better than SNR valu better than igher SNI JR values	speech the NAT mo es but is wor the NAT mo R values and	odel in se at low odel in	
B) 0 5 0 5 5 7 G. 0 5 7 G.	3.027 $3.27$ $2.701$ $3.27$ $2.380$ $3.27$ $2.380$ $3.27$ $2.072$ $3.27$ $1.791$ $2.27$ $1.503$ $2.27$ $2.246$ $3.27$ $3.6$ for prop         Noisy       N	.506       3.686         .394       3.481         .265       3.240         .114       2.982 <b>932</b> 2.708 <b>708</b> 2.409         .153       3.084 <b>oosed and NA</b> VAT       LogFFT	3.720         3.511         3.258         2.975         2.665         2.327         3.076	3.765         3.590         3.380         3.134         2.845         2.513         3.205		the quantum the quantum the quantum the quantum term of the N enhant SNR for the N enhant term of term	ality an NRM mo icing qu values a NRM mo icing int	d intelle odel per ality at nd on a odel per elligibit rse at lo	igibility forms b higher S verage forms b lity at h ower SN	of noisy better than SNR valu better than igher SNI	speech the NAT mo es but is wor the NAT mo R values and	odel in se at low odel in	
3) G. R B)	3.027 $3.027$ $2.701$ $3.027$ $2.701$ $3.027$ $2.380$ $3.027$ $2.380$ $3.027$ $1.791$ $2.072$ $1.791$ $2.072$ $1.503$ $2.072$ $2.246$ $3.027$ $3.0272$ $3.027$ $3.0272$ $3.027$ $1.791$ $2.072$ $1.791$ $2$	.506       3.686         .394       3.481         .265       3.240         .114       2.982 <b>932</b> 2.708 <b>708</b> 2.409         .153       3.084	3.720 3.511 3.258 2.975 2.665 2.327 3.076 <b>X systems in s</b>	3.765 3.590 3.380 3.134 2.845 2.513 3.205 seen noise		the quantum the quantum the quantum the quantum tender of the Nervice of the Nerv	I ality an IRM mo I cing qu Values a IRM mo I cing int Shtly wo	d intelle odel per ality at nd on a odel per elligibit rse at lo	igibility forms b higher S verage forms b lity at h ower SN Sun	of noisy better than SNR valu better than igher SNI JR values	speech the NAT mo es but is wor the NAT me R values and	odel in se at low odel in on avera	٤
) 	3.027 $3.027$ $2.701$ $3.027$ $2.701$ $3.027$ $2.380$ $3.027$ $2.380$ $3.027$ $1.791$ $2.072$ $1.791$ $2.072$ $1.503$ $2.072$ $2.246$ $3.027$ $2.246$ $3.027$ NoisyN $0.961$ $0.026$ $0.926$ $0.026$	.506       3.686         .394       3.481         .265       3.240         .114       2.982 <b>932</b> 2.708 <b>708</b> 2.409         .153       3.084 <b>oosed</b> and NA         VAT       LogFFT         .937 <b>0.981</b>	3.720         3.511         3.258         2.975         2.665         2.327         3.076 <b>T systems in s</b> FFT-MASK         0.974	3.765         3.590         3.380         3.134         2.845         2.513         3.205         seen noise         NRM         0.977		the quantum of the quantum of the NR NR of the	ARM monopole ARM monopole And a strain of the second secon	Id intella odel per ality at and on a odel per celligibil rse at lo	igibility forms b higher S verage forms b lity at h ower SN <b>Sun</b>	of noisy better than SNR valu better than igher SNI JR values <b>nmar</b> d noise ta	speech the NAT mo es but is wor the NAT me R values and <b>Y</b>	odel in se at low odel in on avera	٤
es	3.027 $3.027$ $2.701$ $3.027$ $2.701$ $3.027$ $2.380$ $3.027$ $2.380$ $3.027$ $2.072$ $3.027$ $1.791$ $2.072$ $1.791$ $2.072$ $2.246$ $3.027$ $2.246$ $3.027$ NoisyN $0.961$ $0.026$ $0.926$ $0.026$ $0.872$ $0.026$	.506       3.686         .394       3.481         .265       3.240         .114       2.982         932       2.708         708       2.409         .153       3.084         oosed       and NA         VAT       LogFFT         .937       0.981         .928       0.968	3.720         3.511         3.258         2.975         2.665         2.327         3.076 <b>X</b> systems in s         FFT-MASK         0.974         0.958	3.765         3.590         3.380         3.134         2.845         2.513         3.205         seen noise         NRM         0.977         0.962		the quantum of the quantum of the NR NR of the	ARM monopole ARM monopole And a constant ARM monopole ARM	Id intella odel per ality at and on a odel per celligibil rse at lo ne best a seen no	igibility forms b higher S verage forms b lity at h ower SN <b>Sun</b> all-roun	of noisy better than SNR valu better than igher SNI JR values <b>nmar</b> d noise ta ditions an	speech the NAT mo es but is wor the NAT me R values and <b>Y</b> arget. It outpend in improvis	odel in se at low odel in on avera	٤ t
) 	3.027 $3.027$ $2.701$ $3.027$ $2.701$ $3.027$ $2.380$ $3.027$ $2.380$ $3.027$ $2.072$ $3.027$ $1.791$ $2.072$ $1.791$ $2.072$ $2.246$ $3.027$ $2.246$ $3.027$ $0.961$ $0.0799$ $0.926$ $0.0799$ $0.799$ $0.0799$	.506       3.686         .394       3.481         .265       3.240         .114       2.982 <b>932</b> 2.708 <b>708</b> 2.409         .153       3.084 <b>0osed</b> and NA         NAT       LogFFT         .937 <b>0.981</b> .928 <b>0.968</b> .916 <b>0.947</b>	3.720         3.511         3.258         2.975         2.665         2.327         3.076 <b>X</b> systems in s         FFT-MASK         0.974         0.934	3.765         3.590         3.380         3.134         2.845         2.513         3.205         seen noise         NRM         0.977         0.962         0.941		the quantum of the quantum of the NR NR of the	ARM monopole ARM monopole And a constant ARM monopole ARM	Id intella odel per ality at and on a odel per celligibil rse at lo ne best a seen no	igibility forms b higher S verage forms b lity at h ower SN <b>Sun</b> all-roun	of noisy better than SNR valu better than igher SNI JR values <b>nmar</b> d noise ta ditions an	speech the NAT mo es but is wor the NAT me R values and <b>Y</b>	odel in se at low odel in on avera	٤ t
3) R 3)	3.027 $3.027$ $2.701$ $3.027$ $2.701$ $3.027$ $2.380$ $3.027$ $2.380$ $3.027$ $2.072$ $3.027$ $1.791$ $2.072$ $1.791$ $2.072$ $2.246$ $3.027$ $1.503$ $2.072$ $0.961$ $0.0799$ $0.926$ $0.0799$ $0.799$ $0.0708$ $0.708$ $0.0708$	5063.686.3943.481.2653.240.1142.9829322.7087082.409.1533.084osed and NANATLogFFT.9370.981.9280.968.9160.947.8970.917	3.720         3.511         3.258         2.975         2.665         2.327         3.076 <b>X</b> systems in s         • FFT-MASK         0.974         0.958         0.934         0.899	3.765         3.590         3.380         3.134         2.845         2.513         3.205         seen noise         NRM         0.977         0.962         0.941         0.910		the quantum of the quantum of the NR NR of the	ARM mo acing qu values a VRM mo acing int acing int acin	Id intella odel per ality at and on a odel per celligibil rse at lo ne best a seen no	igibility forms b higher S verage forms b lity at h ower SN <b>Sun</b> all-roun	of noisy better than SNR valu better than igher SNI JR values <b>nmar</b> d noise ta ditions an	speech the NAT mo es but is wor the NAT me R values and <b>Y</b> arget. It outpend in improvis	odel in se at low odel in on avera	٤ t

STOI scores for proposed and NAT systems in seen noise



Babafemi O. Odelowo and David V. Anderson



Center for Signal & Image Processing

- ver, it is
- g both
- ver
- age but

the alues