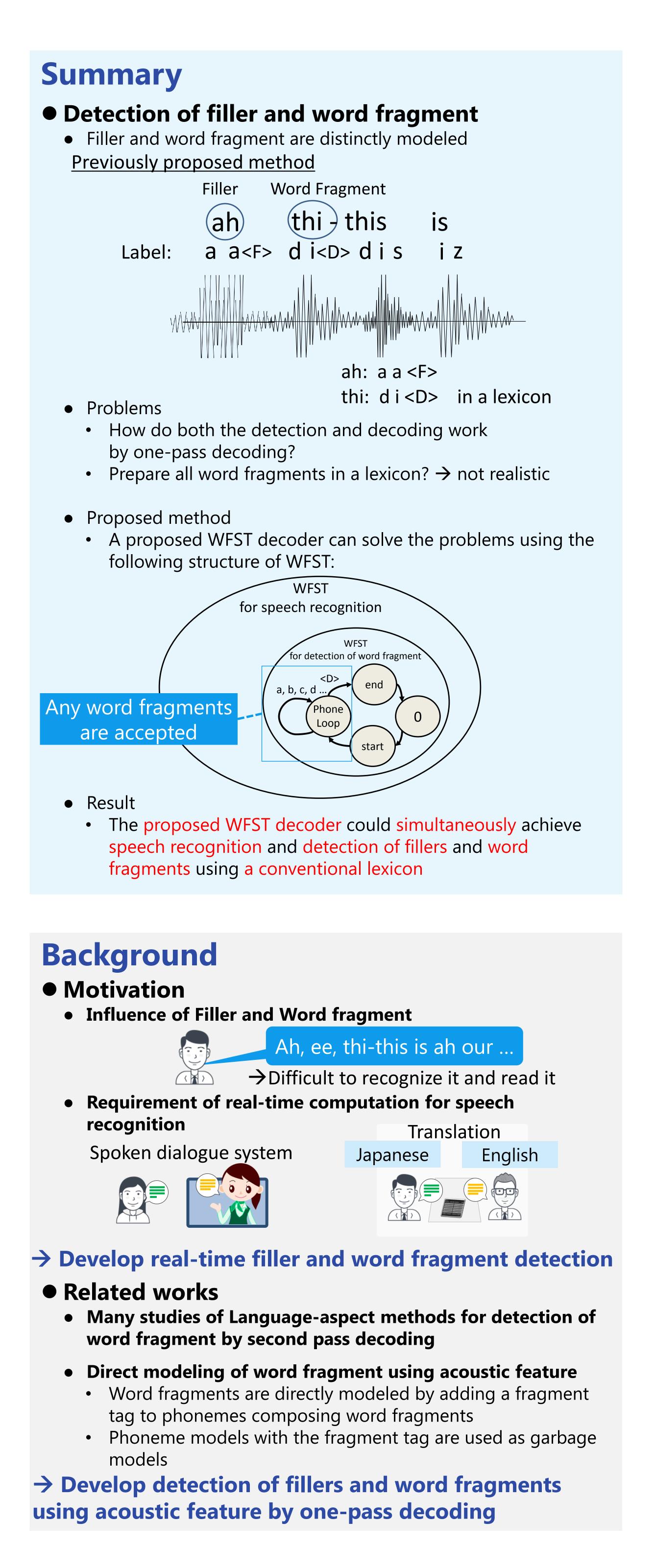
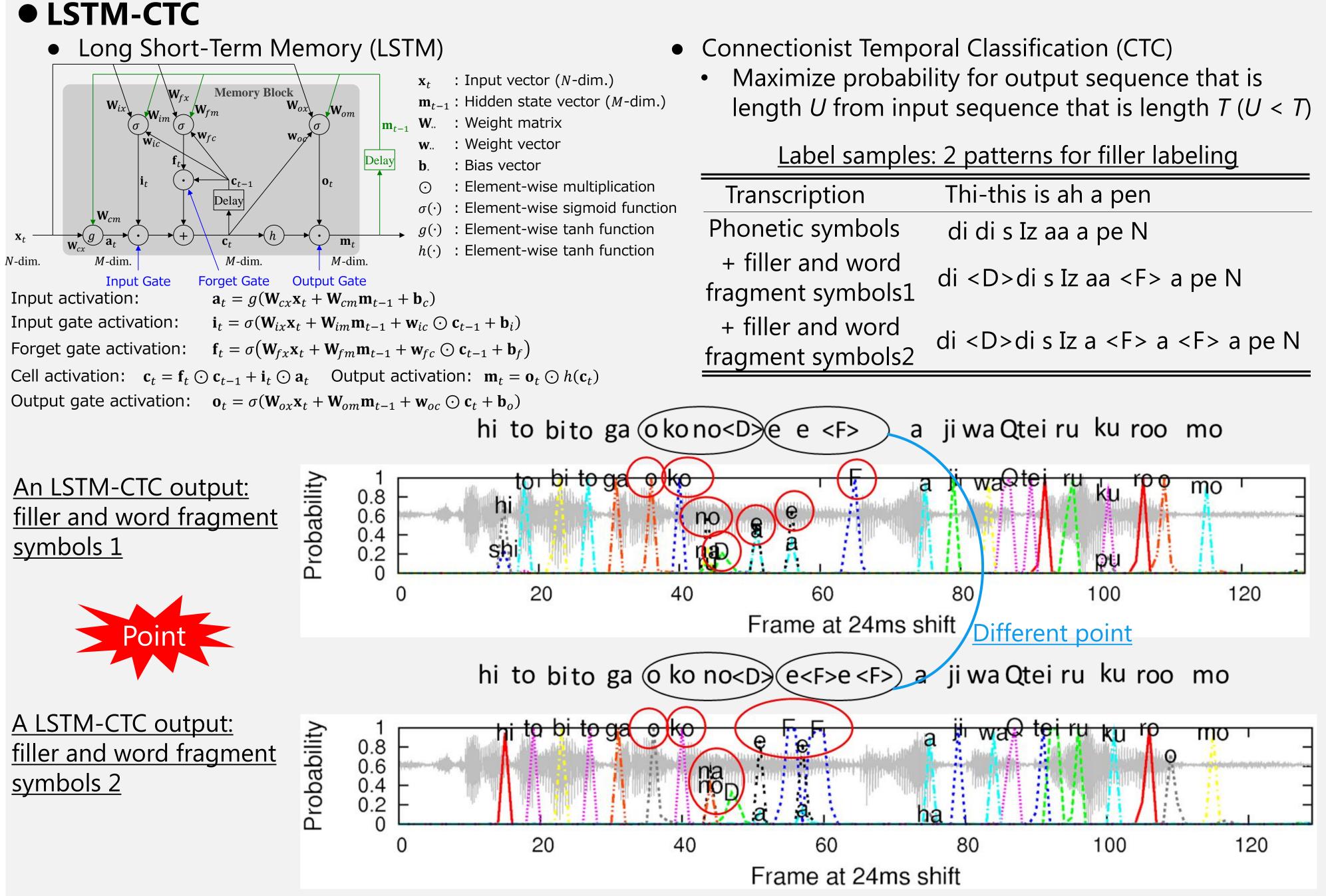


SIMULTANEOUS SPEECH RECOGNITION AND ACOUSTIC EVENT DETECTION USING **AN LSTM-CTC ACOUSTIC MODEL AND A WFST DECODER**



Hiroshi Fujimura^{*1}, Manabu Nagao^{*1*2}, Takashi Masuko Corporate Research & Development Center, Toshiba Corporation

Acoustic Model with detection of fillers and word fragments



Proposed Weighted Finite State Transducer (WFST)

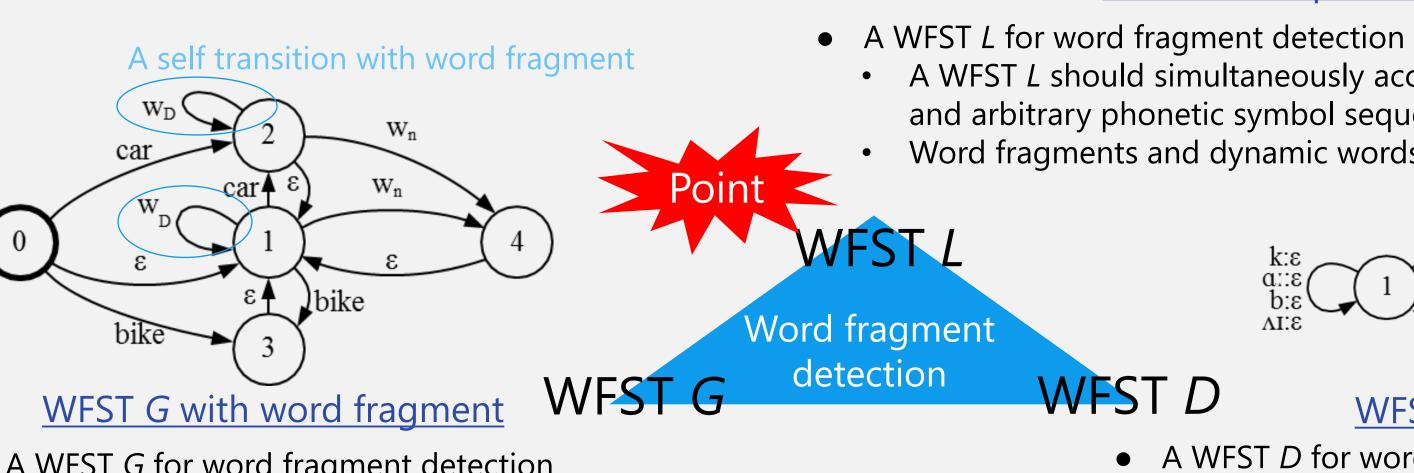
- WFST for general speech recognition
- A WFST for speech recognition consists of *R*, *L* and *G* using LSMT-CTC • *R*: WFST Squashing a label sequence of the acoustic model in a CTC manner
- ex.) "AAAA", " $\Phi \Phi AAA \Phi$ " \rightarrow "A", "A" by squashing (Φ : blank symbol)
- L: WFST of a lexicon, G: WFST of a language model

• WFST for filler detection

• Filler symbol can be detected and deleted by a new WFST *R*

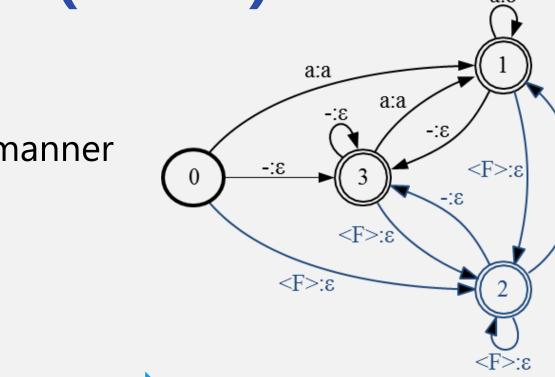
- By checking input symbols on the transitions of word hypotheses, it is possible to detect words as fillers
- Filler confidence score using filler symbol2 (e <F> e <F>) • Confidence score c = f/p,
- where *f*: the number of detected filler symbols *p*: the number of phonetic symbols in a word ex.) ah (a <F> a <F>) c = 1.00,
- third (s a $\langle F \rangle$ a d) c= 0.25

• WFST for word fragment detection

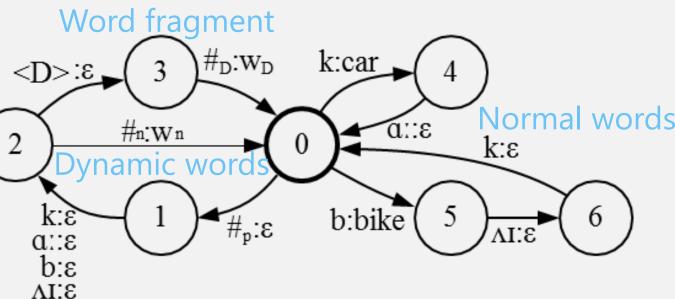


- A WFST *G* for word fragment detection
- A self transition with word fragment w_D is added to a WFST G • A dynamic word w_n is regarded as a normal word
- A combined WFST $RLG = \pi_{\varepsilon}(opt(R \circ opt(proj_{i \to o}(L \circ G))))$

amples	: 2 patterns for filler labeling
٦	Thi-this is ah a pen
bols	di di s Iz aa a pe N
vord bols1	di <d>di s Iz aa <f> a pe N</f></d>
vord bols2	di <d>di s Iz a <f> a <f> a pe N</f></f></d>
a	ji wa Qtei rukuroomo
	i waQtei ru i roo mo i

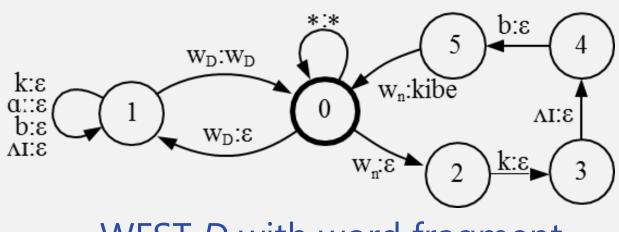


WFST *R* with transitions <u>accepting a filler symbol <F></u>



WFST L with phoneme loops for word fragment

• A WFST *L* should simultaneously accept normal words, dynamic words and arbitrary phonetic symbol sequences terminated by <D> Word fragments and dynamic words share phone loop self-transitions



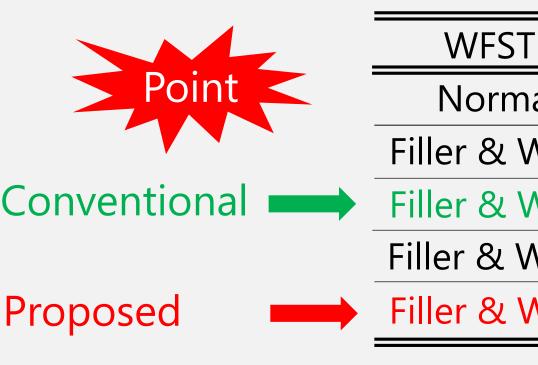
WFST D with word fragment • A WFST *D* for word fragment detection • Convert a phoneme sequences to dynamic words Convert phonetic symbol sequences to w_D

• A combined WFST for decoding $RLGD = RLG \circ D$

•	*1:Equal c	ontributio	on *2: Cu	rrently v	with losh	0			orporat	lon
Experiment	al Setup			_		<u>L311VI-</u>	CTC tra	28-dimen ⁻	tional	
Eva	luation Datasets				Fea	ture		Mel filter		
St	yle Speaker	#Filler #w	ord Fragme	nt Fe	eature cor		on 1+	previous 8	frames	
CSJ testset3 Mor	nologue ⁵ males 5 females	785	169		#LSTN	y		3		
Liaison-meeting Mor		238	36		Memory u			1024		
					ecurrent p	projection		1024→2 Japanese		
<u>Training lab</u>	els for acoustic i				Outpu Trainin		+ fille	r & word f	ragment	
Acoustic Model	"こ, ええと この		ono)"							
Normal model (NAM)		o ko no		_			model	training		
Filler + word fragment detection model (FDAM Filler + word fragment	1) ko <d> e</d>	e to <f> ko</f>	no		N-g Trainin	ram g data	200	4-grar million se from W	entences	
detection model (FDAM		e <f> to <f< td=""><td>> ko no</td><td></td><td>Vacabul</td><td></td><td>+ Tos</td><td>hiba Interr</td><td>nal datas</td><td>set</td></f<></f>	> ko no		Vacabul		+ Tos	hiba Interr	nal datas	set
	<u>Lexicons</u>				Vocabul			200,00	0	
Lexicon							ce = (l)	- c) / c		
Normal lexicon (NLex)		o: ko no, o: e e to		Reference :						
Normal lexicon with		o no, ko: ko,		Detected S						
word fragments (Nlex+D) eeto: e e to				OR S	Section Le	ength of OR	section 1	Length o	f OR Section	on 1
Lexicon with filler and word kono: ko no, ko: ko <d>, fragment symbols 1 (FDLex1) eeto: e e to <f></f></d>				AND S	Section	ngth of AND			f AND Sect	
Lexicon with filler and v fragment symbols 2 (FD	<d>, <f></f></d>						Time	→		
		reated WI					aiculati	<u>ng the to</u>		<u> </u>
	WFST name		oustic Mo	<u> </u>	xicon	Decc	dor	=		
Point <	Normal (NW		NAM		ex+D			=		
	Filler & WF (FD	-	FDAM1		DLex1	Conver Conver				
Conventional	Filler & WF (FD	•	FDAM2		DLex2	Conver				
	Filler & WF (FD		FDAM1		VLex	Propo				
Proposed	Filler & WF (FD	WT4)	FDAM2	Ν	VLex	Propo	_			
					Fil	ler confi	dence	score		
Results				0.9	<u></u>					
ASR performance	<u>e for each WFST</u>	<u>(CER [%])</u>		0.8	* * *					
WFST name CSJ tes	stset3 Liaison-me	eting Ave	<u> </u>	•			····•�	→Precision	(CSJ)	
NWT 10.	.34 15.36	12.8	— 5	0.7		Q		→ Recall (C	SJ)	
FDWT1 10.	.35 14.75	12.5	5	0.6	<u>o</u> <u>o</u> <u>o</u> <u>o</u>		000	··• Precision	(Meeting)	
FDWT2 9.	.83 14.65	12.24	4	0.5		(C)	<u>A</u> <u>A</u> <u>A</u>	⊶ <u>a</u> Recall (M ⊶o F−value (I		
FDWT3 10.	.53 14.82	12.68	3	0.4 —						
FDWT4 10.	.16 14.99	12.57	7		0.1 0.2 0.3 0. Filler conf	4 0.5 0.6 0.7 idence thres				
Filler Detection pe	<u>erformance (F: F-</u>	<u>value)</u>	Point	Word Fi				formance	<u>e (F: F-v</u>	<u>value)</u>
CSJ testset WFST Precision Recall		n-meeting Recall	F	WFST	CSJ Precision	testset3 Recall	F	Liaisor Precision	n-meetin Recall	ig F
NWT 0.77 0.61	0.68 0.79).57 =	NWT	0.22	0.04	0.07	0.00	0.00	0.00
FDWT1 0.80 0.77	0.78 0.87		.69	FDWT1	0.36	0.02	0.04	1.00	0.03	0.05
FDWT2 0.78 0.81	0.79 0.78	0.58 0	0.66	FDWT2	0.42	0.03	0.06	1.00	0.06	0.11
FDWT3 0.75 0.85	0.79 0.72	0.58 0	0.64	FDWT3	0.53	0.25	0.34	0.86	0.33	0.48
FDWT4 0.75 0.85	0.79 0.70	0.54 0	0.61	FDWT4	0.49	0.25	0.34	0.61	0.31	0.41
 The proposed fillers and work The proposed lexicons included 	rd fragment decoder ou	s tperfor	med a	conv	entior	nal de	coder	using		

	1:Equal contrib	ution *2: Cu	irrenti	ly with	n Iosh	J			orporat	lon
Experimental Se	etup						CTC tra	28-diment	tional	
Evaluation	Datasets				Feat	ure		Mel filter		
Style	Speaker #Filler	#Word Fragme	ent	Featu	ire con	catenatio	on 1+	previous 8	frames	
CSJ testset3 Monologue	5 males 785 5 females	169			#LSTM			3		
Liaison-meeting Monologue		36				nit / laye		1024		
				Recu	irrent p	rojectior		1024→2 Japanese		
Training labels for					Output Training		+ fille	r & word f I training s	ragmen	
Acoustic Model "こ,	Training label fo ええと この (ko, eet									
Normal model (NAM)	ko e e to ko no)			Lai	nguage	model	training		
Filler + word fragment detection model (FDAM1)	ko <d> e e to <f< td=""><td>> ko no</td><td></td><td></td><td>N-gr Training</td><td></td><td>200</td><td>4-grar million se</td><td>entences</td><td>5</td></f<></d>	> ko no			N-gr Training		200	4-grar million se	entences	5
Filler + word fragment detection model (FDAM2) ko <	D> e <f> e <f> t</f></f>	o <f> ko no</f>					+ Tos	from Wo hiba Interr		set
Lexic				ocabula	ary size		200,00	0	_	
	nciation				tolerar	nce = (l	- c) / c			
Normal lexicon (NLex)	Normal lexicon (NLex) kono: ko no,				ion					-
Normal lexicon with	eeto: e e to kono: ko no, ko		Detect	ed Secti	ion					
word fragments (Nlex+D)	\mathbf{N}	(OR Secti							
Lexicon with filler and word	exicon with filler and word kono: ko no, ko: ko <d>,</d>				Ler ion	ngth of OR	section 1	Length o	f OR Section	on 1
fragment symbols 1 (FDLex1) Lexicon with filler and word					Leng	gth of AND) Section c	Length o	f AND Sec	→
fragment symbols 2 (FDLex2)		<u>The</u>	<u>e metr</u>	<u>ic for c</u>	<u>alculati</u>	<u>ng the to</u>	Time Dierance			
	Created	l WFSTs ("W	<u>′F″: W</u>	ord Fi	ragme	<u>ent)</u>				
	FST name	Acoustic Mc	del	Lexic	on	Deco	oder	_		
Point No	ormal (NWT)	NAM		Nlex	+D	Conver	ntional			
Filler	& WF (FDWT1)	FDAM1		FDLe	ex1	Conver	ntional			
	& WF (FDWT2)	FDAM2		FDLe		Conver				
	& WF (FDWT3)	FDAM1		NLe		Propo	_			
Proposed Filler	& WF (FDWT4)	FDAM2	•	NLe	2X	Prop	osed	_		
Results					<u>Fill</u>	<u>er conf</u>	<u>idence</u>	<u>score</u>		
			0.9)	* * *	_				
ASR performance for ea	ich WFST (CER [%])	0.8	3						
	5	Ave.	0.7					← Precision	12	
NWT 10.34		12.85	0.6	•·····•·····•·····•·····•·····•·····•····•····•····•····•····•····•····•····•····•····•····•····•····•····•····•····•····•····•····•····•···	. @ @ @ .	<u>0</u> 0	<u>o</u> o	→ Recall (CS	CSJ)	
FDWT1 10.35		12.55		AA	··Δ·····Δ······Δ··			⊶⊶ Precision ⊶⊶ Recall (Me	eeting)	
FDWT2 9.83		12.24	0.5	5		<u>`A</u> A.	·····A·····A	·· <mark>o</mark> · F−value (I	Meeting)	
FDWT3 10.53 FDW/T4 10.16		L2.68 12.57	0.4		020304	050607	0.8 0.9 1.0			
FDWT4 10.16	14.33					dence three				
Filler Detection performa		> Point	Word	d Frag			ion per	formance	-	-
CSJ testset3 WFST Precision Recall F	Liaison-meet Precision Recal	0	WFS	T Pre	CSJ 1 cision	testset3 Recall	F	Liaisor Precision	n-meetir Recall	ng F
NWT 0.77 0.61 0.68	0.79 0.45	0.57	NW	/T 0	.22	0.04	0.07	0.00	0.00	0.00
FDWT1 0.80 0.77 0.78	0.87 0.57	0.69	FDW	Τ <u>1</u> 0	.36	0.02	0.04	1.00	0.03	0.05
FDWT2 0.78 0.81 0.79	0.78 0.58	0.66	FDW	′T2 0	.42	0.03	0.06	1.00	0.06	0.11
FDWT3 0.75 0.85 0.79	0.72 0.58	0.64	FDW		0.53	0.25	0.34	0.86	0.33	0.48
FDWT4 0.75 0.85 0.79	0.70 0.54	0.61	FDW	/T4 0	.49	0.25	0.34	0.61	0.31	0.41
 The proposed deco fillers and word fra The proposed deco lexicons including 	gments der outperf	formed a	n cor	nven	ntion	al de	coder	using		

	*1:Equal contribu	ition *2: Cu	irrently	/ with losh	LSTM-			orporat	lon
Experimental S	etup						28-diment	ional	
Evaluation	n Datasets			Feat	ture		Mel filter		
Style	Speaker #Filler	#Word Fragme	ent	Feature cor		on 1+	previous 8	frames	
CSJ testset3 Monologue	5 males 5 females 785	169		#LSTN			3		
Liaison-meeting Monologue		36	·	#Memory u	-		1024 1024→2		
Training labels for				Recurrent p Outpu		126	Japanese r & word f	moras	t
Acoustic Model	Training label fo ええと この (ko, eeto		-	Trainin	g data	CS.	J training s	et 580h	
Normal model (NAM)	ko e e to ko no		:	La	nguage	mode	training		
Filler + word fragment detection model (FDAM1)	ko <d> e e to <f></f></d>	• ko no	-	N-g Trainin		200	4-grar million se	entences	5
Filler + word fragment detection model (FDAM2) ko	<d> e <f> e <f> to</f></f></d>	o <f> ko no</f>	-			+ Tos	from Wo hiba Interr	nal datas	set
<u>Lexi</u>	<u>cons</u>		=	Vocabul	ary size		200,00	0	
Lexicon W	Lexicon Word and the pronunciation					ce = (l	- c) / c	1	
Normal lexicon (NLex)	kana: ka na								
Normal lexicon with word fragments (Nlex+D)	Normal lexicon withkono: ko no, ko: ko,word fragments (Nlex+D)eeto: e e to			R Section					
Lexicon with filler and word kono: ko no, ko: ko <d>,</d>				Le D Section	ength of OR	section 1	Length o	f OR Section	on 1
fragment symbols 1 (FDLex1) eeto: e e to <f></f>					igth of AND	Section c	Length o	f AND Sec	tion c
Lexicon with filler and word fragment symbols 2 (FDLex2)	kono: ko no, ko: k eeto: e <f> e <f></f></f>	-		<u>The met</u>	<u>ric for ca</u>	alculati	<u>ng the to</u>	Time Dierance	
	Created	WFSTs ("W	′F″: Wo	ord Fragme	<u>ent)</u>				
M	/FST name	Acoustic Mc	del [exicon	Deco	oder	_		
Point N	ormal (NWT)	NAM		Nlex+D	Conver	ntional			
	& WF (FDWT1)	FDAM1		FDLex1	Conver	ntional			
	& WF (FDWT2)	FDAM2		FDLex2	Conver				
	& WF (FDWT3)	FDAM1		NLex	Propo	_			
Proposed Filler	& WF (FDWT4)	FDAM2		NLex	Propo	osed	_		
Results				<u>Fil</u>	<u>ler confi</u>	<u>dence</u>	<u>score</u>		
		/ 1)	0.9		-				
ASR performance for e			0.8			····•			
		Ave.	0.7				♣Precision ♣Recall (CS)	72	
NWT 10.34 FDWT1 10.35		2.85	0.6	<u></u>		000	←F−value(··• Precision	CSJ)	
FDWT2 9.83		2.25	0.5	<u>۸</u> <u>۸</u> <u>۸</u> <u>۸</u>		····· <u>A</u> ······A	·· <u>A</u> ··Recall (Me	eeting)	
FDWT3 10.53		2.68	0.5				······ F - value (I	vieeting)	
FDWT4 10.16		2.57	0.4 (0.0 0.1 0.2 0.3 0. Fillerconf	4 0.5 0.6 0.7 idence thres				
Filler Detection perform	ance (F: F-value)	Point	Word	Fragment	Detecti	on per	formance	<u>e (F: F-v</u>	/alue)
CSJ testset3 WFST Precision Recall F	Liaison-meeti Precision Recall	5	WFS	CSJ Precision	testset3 Recall	F	Liaisor Precision	n-meetir Recall	ng F
NWT 0.77 0.61 0.68	0.79 0.45	0.57	NW	T 0.22	0.04	0.07	0.00	0.00	0.00
FDWT1 0.80 0.77 0.78	0.87 0.57	0.69	FDW	F1 0.36	0.02	0.04	1.00	0.03	0.05
FDWT2 0.78 0.81 0.79	0.78 0.58	0.66	FDW	T2 0.42	0.03	0.06	1.00	0.06	0.11
FDWT3 0.75 0.85 0.79	0.72 0.58	0.64	FDW	F3 0.53	0.25	0.34	0.86	0.33	0.48
FDWT4 0.75 0.85 0.79	0.70 0.54	0.61	FDW	T4 0.49	0.25	0.34	0.61	0.31	0.41
 The proposed dece fillers and word fra The proposed dece lexicons including 	agments oder outperf	ormed a	a con	ventior	nal deo	codei	using		



	*1:Equal co	Jillibuli	011 Z. Cl	unentig	y with iOS	J	CTC tra		прогас	.1011
Experiment	-							28-diment	ional	
<u>Eva</u>	aluation Datasets					iture		Mel filter		
St	yle Speaker a	Filler #\	Nord Fragm	ent	Feature co		on 1+	previous 8 ר	trames	
CSJ testset3 Mor	nologue ^{5 males} 5 females	785	169		#LSTN #Memory (A layer	r	3 1024		
Liaison-meeting Mor	nologue 1 male	238	36		Recurrent			1024→2		
Training lab	pels for acoustic r	nodels				ut class	126	Japanese & word f	moras	 t
Acoustic Model	Training "こ, ええと この (kono)"	:	Trainir	ng data	CS.	training s	et 580h	
Normal model (NAM)	ko e e t	o ko no			La	anguage	model	training		
Filler + word fragment detection model (FDAM	11) KO <d> e e</d>	to <f> k</f>	0 00			gram ng data	200	4-grar million se	entences	 ;
Filler + word fragment detection model (FDAM		e <f> to <</f>	<f> ko no</f>	-			+ Tos	from We hiba Interr	nal datas	set
	<u>Lexicons</u>			:	Vocabu	lary size		200,00	0	
Lexicon	Word and the	•	ation	Referen	ce Section	tolerar.	nce = (l →	- C) / C		-
Normal lexicon (NLex)		: ko no, : e e to			ed Section					
Normal lexicon with word fragments (Nlex+		no, ko: ko : e e to	Э,	0	R Section					
Lexicon with filler and word kono: ko no, ko: ko <d>,</d>				AN	L D Section	ength of OR	section 1	Length of	f OR Section	on 1
fragment symbols 1 (FDLex1) eeto: e e to <f></f>					Le	ngth of AND	Section c	Length of	f AND Sec	•
Lexicon with filler and fragment symbols 2 (FC		-	-	:	<u>The met</u>	<u>tric for c</u>	alculati	ng the to	Time lerance	
	<u></u>	reated W	/FSTs ("V	<u>/F": Wo</u>	ord Fragm	<u>ent)</u>				
	WFST name	A	coustic Mo	odel	Lexicon	Deco	oder	_		
Point <	Normal (NW	T)	NAM		Nlex+D	Conver	ntional			
	Filler & WF (FD)	•	FDAM:		FDLex1	Conver	ntional			
Conventional	Filler & WF (FD)	-	FDAM		FDLex2	Conver				
	Filler & WF (FD)				NLex	Propo	_			
Proposed	Filler & WF (FD)	///////////////////////////////////////	FDAM2		NLex	Prop	osed	_		
Results				0.0	<u>Fi</u>	<u>ller conf</u>	idence	<u>score</u>		
ASP portormance	o for oach M/EST (١	0.9		*				
	e for each WFST			0.8			·····			
WFST name CSJ tes NWT 10	stset3 Liaison-mee .34 15.36	ting Av 12.		0.7	۵	A		→Precision →Recall (CS)	12	
	.35 14.75	12.		0.6	000		·····0·····0	- F−value (0 		
	.83 14.65	12.		0.5	<u>۸</u> <u>۸</u> <u>۸</u>		<u>A</u> <u>A</u> <u>A</u>	<mark>⊶₄</mark> - Recall (Me ⊶⊛- F−value (N	eeting)	
	.53 14.82	12.0							nooting/	
FDWT4 10	.16 14.99	12.	57	0.4	0.0 0.1 0.2 0.3 0	0.4 0.5 0.6 0.7 fidence thre				
Filler Detection pe	erformance (F: F-v	<u>/alue)</u>	Point	Word	Fragmen			formance	<u>e (F: F-v</u>	<u>/alue)</u>
CSJ testset WFST Precision Recall		n-meeting Recall		WFS ⁻	CSJ T Precision	testset3 Recall	F	Liaison Precision	-meetin Recall	ng F
NWT 0.77 0.61	0.68 0.79		0.57	 NW [:]		0.04	г 0.07	0.00	0.00	۲ 0.00
FDWT1 0.80 0.77	0.78 0.87		0.69	FDW	•	0.04	0.07	1.00	0.00	0.00
FDWT2 0.78 0.81	0.79 0.78		0.66	FDW		0.03	0.06	1.00	0.06	0.11
FDWT3 0.75 0.85	0.79 0.72		0.64	FDW		0.25	0.34	0.86	0.33	0.48
FDWT4 0.75 0.85	0.79 0.70	0.54	0.61	FDW ⁻	T4 0.49	0.25	0.34	0.61	0.31	0.41
 The proposed fillers and work The proposed lexicons included 	rd fragments decoder out	s tperfo	rmed a	a con	ventio	nal de	coder	using		

	*1:Equal co	Shtributi	on ~2: Cu	irrenuy	with iosr	0	CTC tra		orporat	.ion
Experimenta	-			=				 28-dimen	tional	
<u>Eval</u>	uation Datasets					ture		Mel filter		
Sty		#Filler #v	Nord Fragme	ent –			on I+	previous 8	3 frames	
CSJ testset3 Mono	ologue ^{5 males} 5 females	785	169			1 layer	r	3 1024		
Liaison-meeting Mond	ologue 1 male	238	36		[#] Memory ι Recurrent μ			1024→2		
Training labe	els for acoustic r	nodels		I	•	t class	126	Japanese r & word	moras	 t
Acoustic Model	Training "こ, ええと この	label for	(ODO)"		Trainin	g data	CS.	I training	set 580h	
Normal model (NAM)	ko e e t			_	<u>La</u>	inguage	mode	training	_	
Filler + word fragment detection model (FDAM1	.) KO < D> e e	e to <f> k</f>	0 NO		N-g Trainin	ram g data	200	4-gra million so	entences	5
Filler + word fragment detection model (FDAM2		e <f> to <</f>	<f> ko no</f>				+ Tos	from W hiba Inter		set
	Lexicons			_	Vocabu	ary size		200,00	0	
Lexicon							ce = (l	- c) / c		
Normal lexicon (NLex)		: ko no,		Reference	e Section					-
		: e e to		Detected	Section					
Normal lexicon with kono: ko no, ko: ko, word fragments (Nlex+D) eeto: e e to				OR	Section					
Lexicon with filler and w	vord kono: ko r		<d>,</d>	AND	Le Section	ength of OR	section 1	Length c		on 1
fragment symbols 1 (FDL			ngth of AND	Section c	Length c	of AND Sec	tion c			
Lexicon with filler and w fragment symbols 2 (FDI	<d>, o <f></f></d>		The met	<u>ric for c</u>	alculati	ng the to	Time Dierance			
	<u>C</u>	reated W	/FSTs ("W	<u>'F": Wor</u>	<u>rd Fragm</u>	<u>ent)</u>				
	WFST name	Ac	coustic Mc	del Le	exicon	Decc	oder			
Point <	Normal (NW	/Τ)	NAM	Ν	llex+D	Conver	ntional	_		
	Filler & WF (FD	WT1)	FDAM1	. F	DLex1	Conver	ntional			
Conventional	Filler & WF (FD	WT2)	FDAM2	. F	DLex2	Conver	ntional			
-	Filler & WF (FD)	WT3)	FDAM1		NLex	Propo	osed			
Proposed	Filler & WF (FD	WT4)	FDAM2		NLex	Propo	osed			
Results					<u>Fil</u>	<u>ler confi</u>	idence	<u>score</u>		
nesurs				0.9						
ASR performance	for each WFST	(CER [%]))	0.8						
WFST name CSJ test	tset3 Liaison-mee	eting Ave	e	0.7 +		0	·····�	→Precision	(CSJ)	
NWT 10.3	34 15.36	12.8	85	*	·····•			- → -Recall (C - → -F-value (
FDWT1 10.3	35 14.75	12.5	55	0.6 [©]	······&·····A·····A·····A·····		·····Q·····Q·····Q	·· <mark>o</mark> · Precision ·· <u>o</u> · Recall (M	89 <u>77</u> 831	
FDWT2 9.8	33 14.65	12.2	24	0.5			<u>A</u> <u>A</u> <u>A</u>	⊷⊛· F−value (
FDWT3 10.5		12.6	58	0.4 +						
FDWT4 10.1	14.99	12.5	57	0.0	0.1 0.2 0.3 0. Fillerconf	4 0.5 0.6 0.7 fidence thres				
Filler Detection per	rformance (F: F-	value)	Point	Word F	- ragment	<u>: Detecti</u>	<u>on per</u>	formanc	<u>e (F: F-v</u>	<u>/alue)</u>
CSJ testset3 WFST Precision Recall	Liaiso F Precision	n-meeting Recall	F	WFST	CSJ Precision	testset3 Recall	F	Liaiso Precision	n-meetin Recall	ng F
NWT 0.77 0.61	0.68 0.79		0.57	NWT		0.04	0.07	0.00	0.00	0.00
FDWT1 0.80 0.77	0.78 0.87	0.57	0.69	FDWT	1 0.36	0.02	0.04	1.00	0.03	0.05
FDWT2 0.78 0.81	0.79 0.78	0.58	0.66	FDWT	2 0.42	0.03	0.06	1.00	0.06	0.11
FDWT3 0.75 0.85	0.79 0.72	0.58	0.64	FDWT	3 0.53	0.25	0.34	0.86	0.33	0.48
FDWT4 0.75 0.85	0.79 0.70	0.54	0.61	FDWT	4 0.49	0.25	0.34	0.61	0.31	0.41
 The proposed fillers and wor The proposed lexicons includ 	d fragments decoder out	s t <mark>perfo</mark> i	rmed a		ventior	nal de	coder	using		

*1:Equal contribution *2: Currently with Toshiba Digital Solutions Corporation

lexicons including word fragments in word fragment detection