



NATIONAL RESEARCH
UNIVERSITY

AMD HSE

ASR. DOMAIN LANGUAGE MODEL IN END-TO-END

Our experience

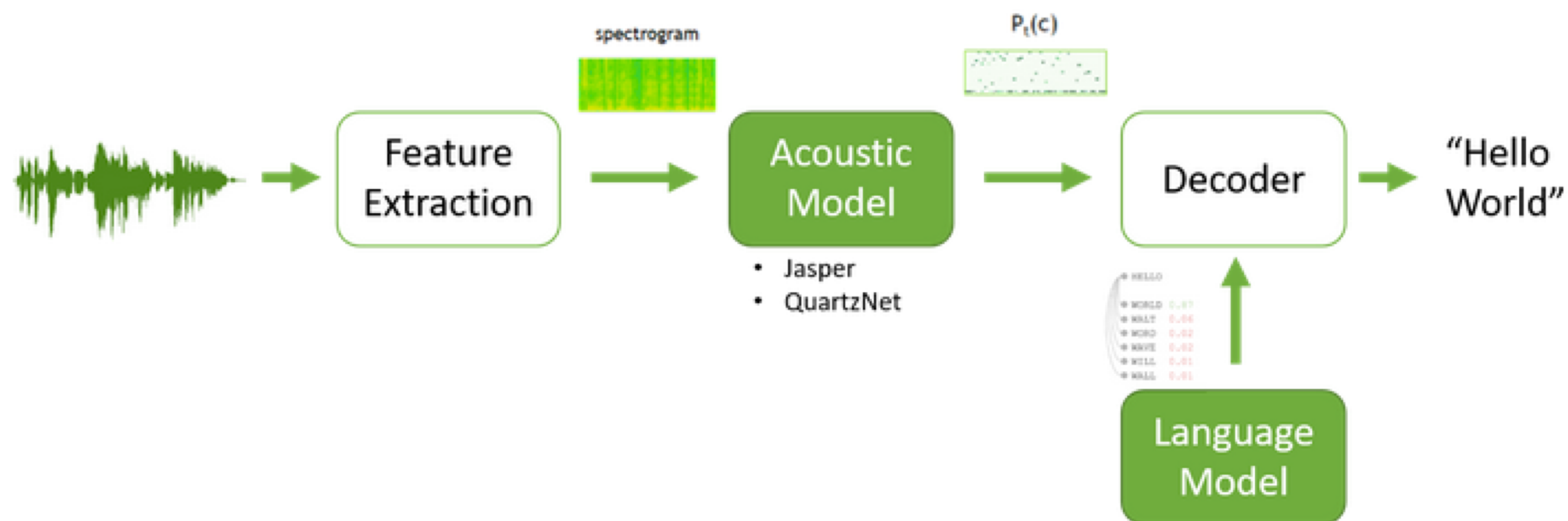
Nizhniy Novgorod, 2020



DOMAIN LANGUAGE MODEL IN END-TO-END

AMD HSE

ASR pipeline. Just to remind



Decoding:

- Argmax
- Beamsearch
- Beamsearch + Language model scoring



DOMAIN LANGUAGE MODEL IN END-TO-END

AMD HSE

Problem:

- Inaccurate recognition due to accent

GT: make a call

Text: make a cool

- Inaccurate recognition due to gap in data

GT: turn off all weekday alarms

Text: turn off all weak gay alarms

- Inaccurate recognition due to specific customized cases

GT: call freddie

Text: call freddy



DOMAIN LANGUAGE MODEL IN END-TO-END

AMD HSE

Contextual decoding. Domains. Domain LM.



Domain unigrams:

1. \$CONTACT
2. \$ADDRESS
3. \$APPLICATION
4. ...

Biasing.

Please call **William**.

Call **mum** now.

Call **Ella** via **WhatsApp**

Call **911**

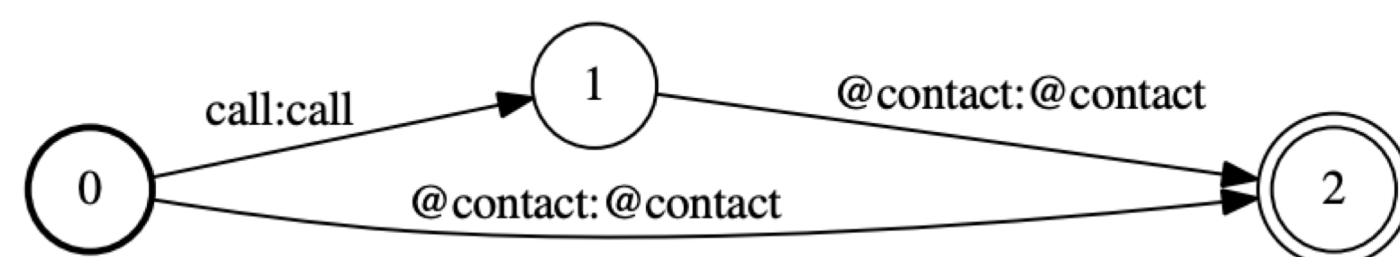


Please call **\$contact**.

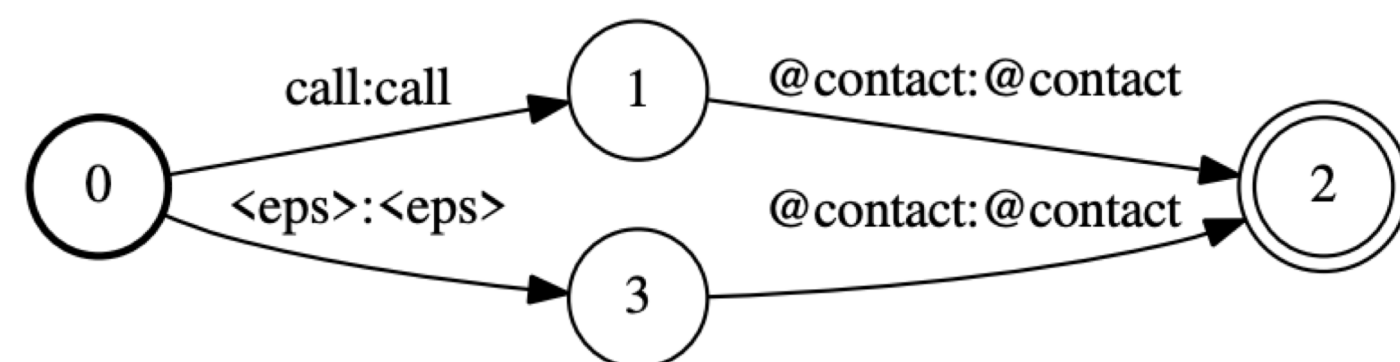
Call **\$contact** now.

Call **\$contact** via **\$app**

Call **\$number**



(a)





Papers

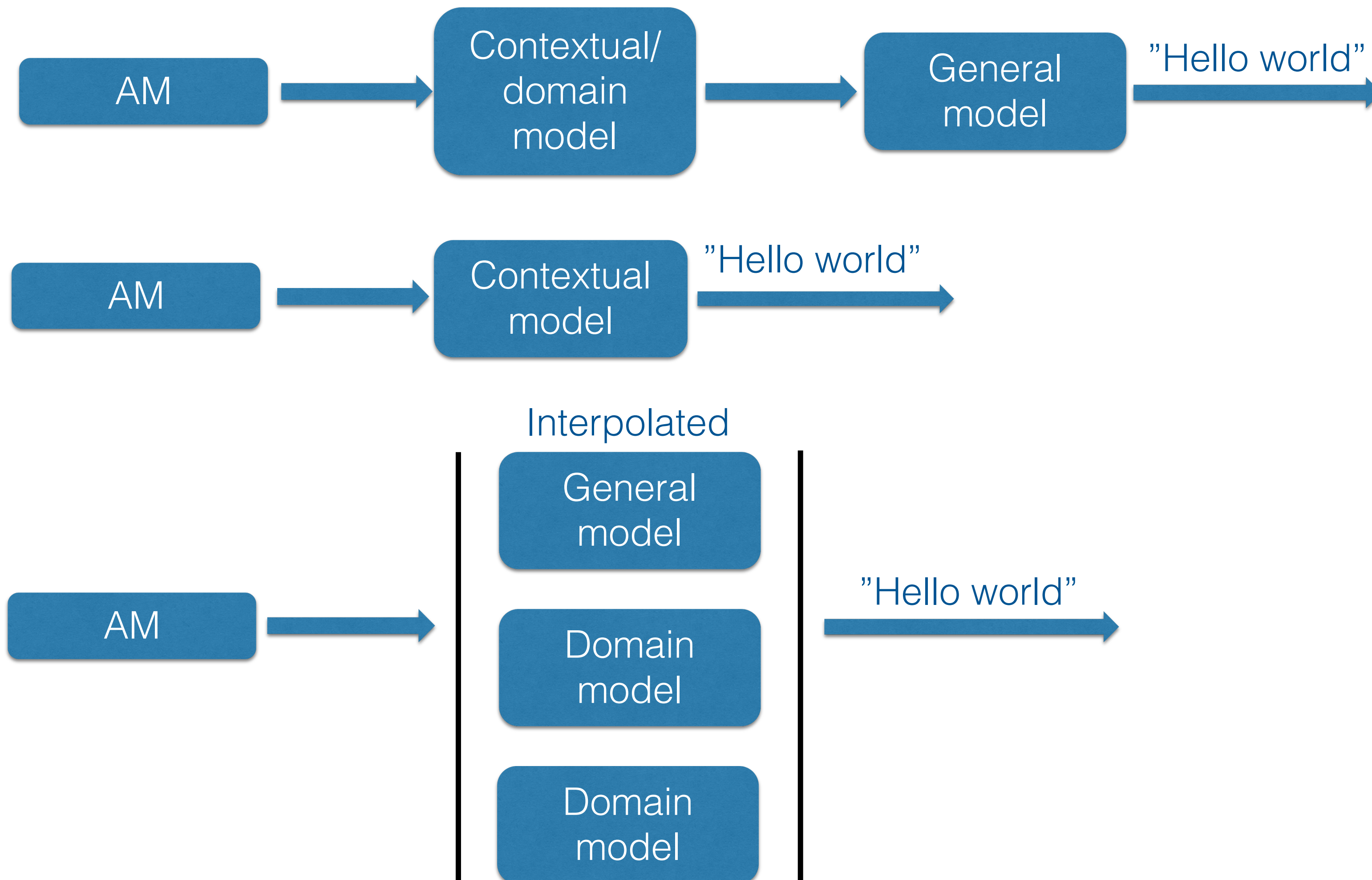
- He Y. et al. Streaming end-to-end speech recognition for mobile devices //ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2019. – C. 6381-6385.
- Williams I. et al. Contextual Speech Recognition in End-to-end Neural Network Systems Using Beam Search //Interspeech. – 2018. – C. 2227-2231.
- McGraw I. et al. Personalized speech recognition on mobile devices //2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2016. – C. 5955-5959.
- Aleksic P. et al. Improved recognition of contact names in voice commands //2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2015. – C. 5172-5175.
- Hall K. et al. Composition-based on-the-fly rescoring for salient n-gram biasing. – 2015.



DOMAIN LANGUAGE MODEL IN END-TO-END

AMD HSE

Existed approaches

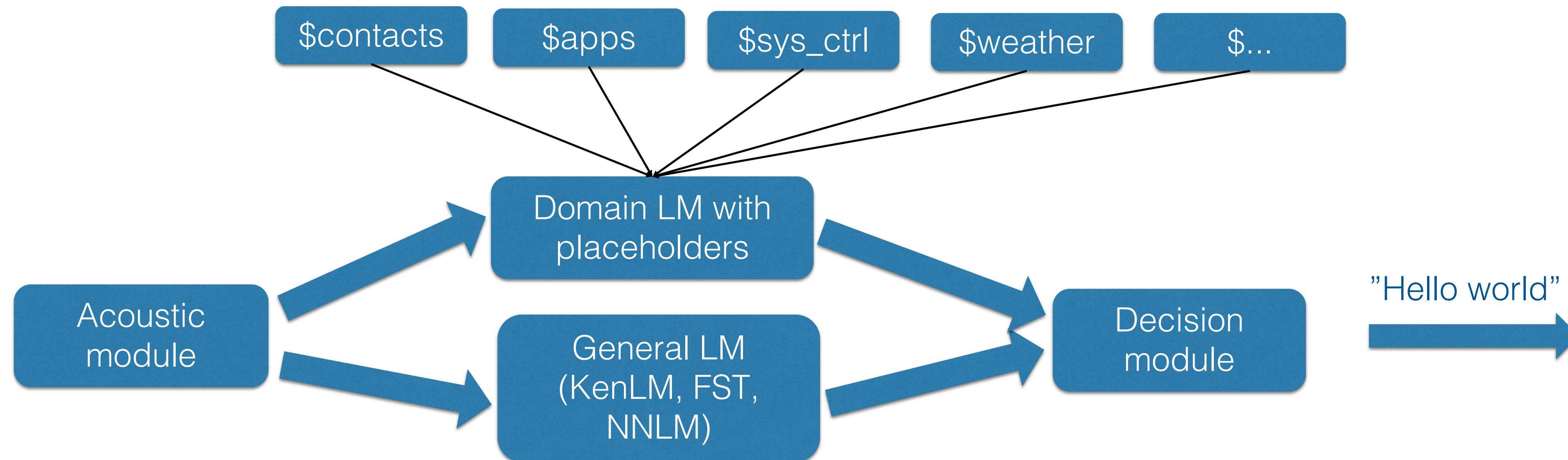




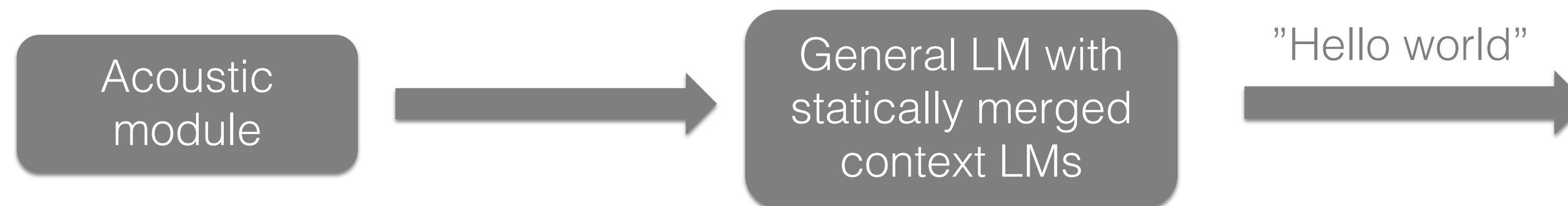
DOMAIN LANGUAGE MODEL IN END-TO-END

AMD HSE

Dynamic parallel scoring



Off-line merged





DOMAIN LANGUAGE MODEL IN END-TO-END

AMD HSE

Experiments:

Tests	# of audio recording	domains includes	Example
Calls	457	names (\$CONTACT) Phone numbers (\$DIGIT) application (\$APP)	call-my-mom call-0845400033 make-a-call-mary-the-second-one make-a-call-using-twitter
Alarms	275	numbers (\$DIGIT) time (\$TIME)	is-my-wakeup-alarm-set what-are-my-weekend-alarms set-an-alarm-for-9-am-the-day-after-tomorrow switch-off-alarm



DOMAIN LANGUAGE MODEL IN END-TO-END

AMD HSE

Tests And Results:

Model	CALLS (WER)	ALARMS (WER)
ArgMax	26	15.4
General LM (Baseline)	25.54	14.63
Off-line merged LM	16.0	9.5
Domain based LM + General LM	10.4	9.8

Analysis:

Dataset	Domain	Off-line
Calls	call-ava.wav "call ava " call-01223454541.wav "call one two two three four five four five four one "	call-ava.wav " over " make-a-call-to-vodafone.wav " make a call to vote phone " make-a-call-to-vodafone.wav " make a cause of vodafone " "make a cool one three seven one one two three four five six seven eight " call-voicemail.wav " cool voice mail "
Alarms	set-alarm-for-tomorrow.wav " set a long for tomorrow " how-many-morning-alarms-do-i-have.wav " how many mourning alarms do i have " are-there-alarms-set-for-tomorrow.wav " other alarms set for tomorrow "	list-my-alarms.wav " please my alarms " unmute-alarm.wav " un metal " increase-alarm-volume.wav " increase along volume " are-there-alarms-set-for-tomorrow.wav " other alarm set for tomorrow "



NATIONAL RESEARCH
UNIVERSITY