BASELINE SPEECH RECOGNITION SYSTEM FOR WEATHER DOMAIN

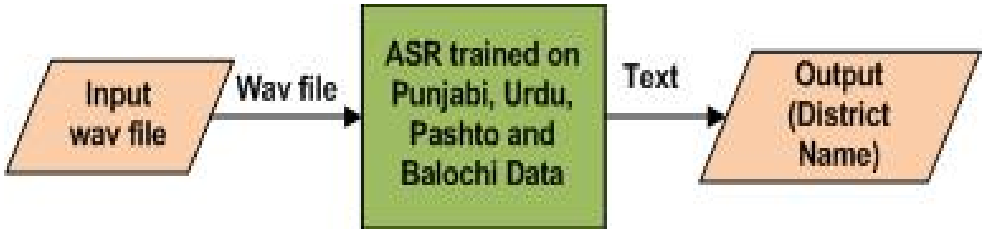
14th April, 2015 | Center for Language Engineering (CLE)

DIALOG FLOW

سسٹم : سی ایل ای موسم کا حال میں خوش آمدید۔
سسٹم: آپ پاکستان کے کس شہر کے موسم کا حال جاننا چاہتے ہیں۔ بیپ کے بعد اس شہر کا نام بتائیے
کالر: لاہور
سسٹم : براہِ مہربانی انتظار فرمائیے گا۔
سسٹم: اگلے ۲۴ گھنٹوں کے دوران لاہور کا درجہ حرارت بیس ڈگری سینٹی گریڈ رہے گا۔ اور وقفے وقفے کے ساتھ ہلکی بارش کا امکان ہے۔
سسٹم: کیا آپ پاکستان کے کسی اور شہر کے موسم کا حال جاننا چاہتے ہیں؟ ہاں یا نہیں میں جواب دیجیے
کالر: نہیں
سسٹم: آپکا کال کرنے کا شکریہ۔

BASELINE ACCENT INDEPENDENT SPEECH RECOGNITION SYSTEM FOR WEATHER DOMAIN

Architecture Diagram

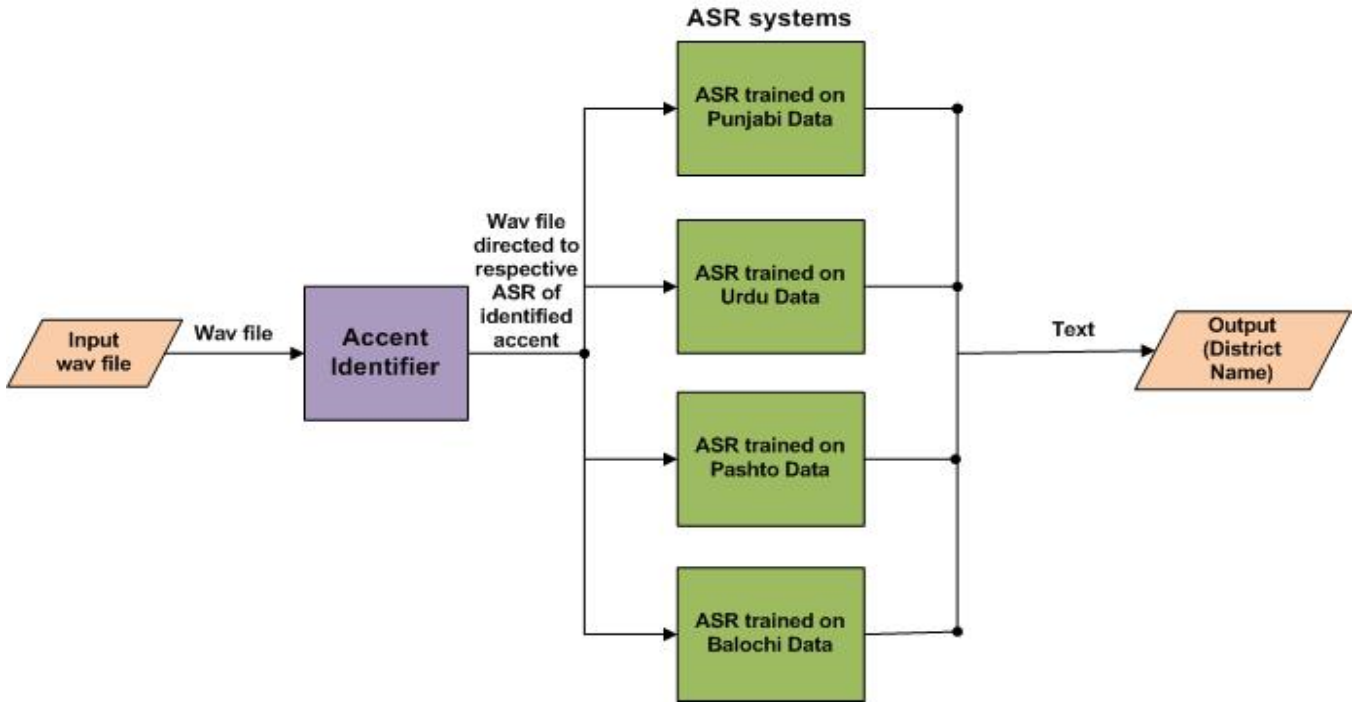


Offline word ASR Results

Accent	Vocabulary size	Training Utterances	Testing Utterances	Accuracy (%age)
Punjabi, Urdu, Pashto, Balochi	139	31802	10216	91.87

BASELINE ACCENT DEPENDENT SPEECH RECOGNITION SYSTEM FOR WEATHER DOMAIN

Architecture Diagram



OFFLINE RESULTS

Accuracy of Accent Identifier

Accent	Training Files	Testing Files	Correctly identified	Accuracy
Balochi	3670	1995	1439	72.13%
Pashto	3670	1771	839	47.37%
Punjabi	3670	988	464	46.96%
Urdu	3670	4341	3234	74.49%
All Accents		9095	5976	65.71%

Accuracy of word ASR system

Accent	Vocabulary Size	Training Utterances	Testing Utterances	Accuracy (%age)
Punjabi	139	3670	988	91.29
Urdu	139	17080	4341	95.09
Pashto	139	6781	1771	90.06
Balochi	139	4271	1995	90.82
			Overall AD ASR System Accuracy	92.76

FIELD TESTING

The purpose of conducting field-testing of ASR system is to evaluate system performance in the scenarios and places where the system is intended to be used, and hence get the feel of how system will perform in real-world scenarios.

Offline Testing	Field Testing
Silence is precisely cut from speech manually	Silence is cut from speech automatically using Voice Activity Detector outlined in (Rabiner & Sambur, February 1975)
Noisy files are separated from test file manually	Noisy files are part of test files
Out-Of-Vocabulary (OOV) and mispronounced words are also removed from the testing data.	Out-Of-Vocabulary (OOV) and mispronounced words are removed using methodology given in (Irtza, Anwar, & Hussain, 2014).

SELECTED NOISE SCENARIOS AND DEMOGRAPHICS

Based on the amount of noise present in the surroundings, from very quiet environment to very loud, different places selected were

- Labs
- offices, classrooms
- campus-parking space
- open-fields (campus lawns)
- cafeteria
- bus-stand and roads within the campus

Demographics include:

- Technical people involved with the project
- Technical people not involved with the project
- Non-technical staff, students, car and rickshaw drivers, shopkeepers and waiters of the cafeteria

FIELD ACCURACY OF DIALOG SYSTEM WITH ACCENT INDEPENDENT ASR

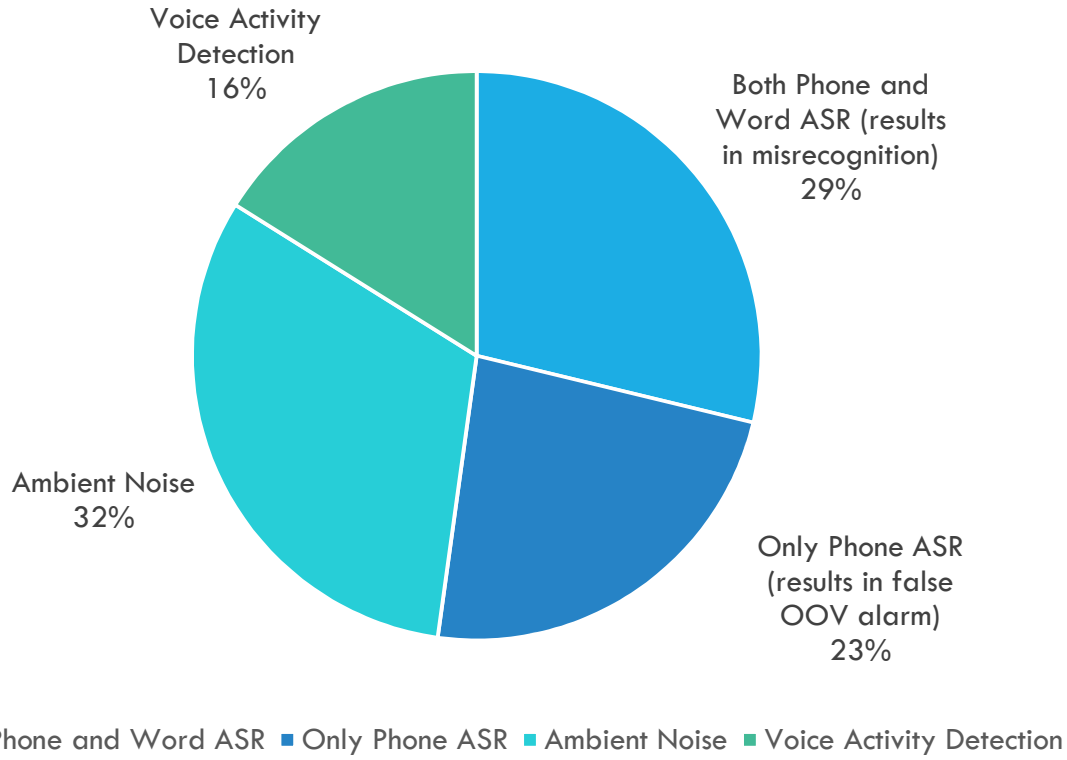
The accuracy of complete dialog system is measured in terms of the response it generates and how it handles the error cases.

Complete end to end Dialog accuracy:

No. of Speakers	Total Test Files	Correct System Response		Incorrect System Response	Overall System Accuracy
		In-vocabulary word correctly decoded	OOV or Multiple words correctly identified	In-vocabulary words misrecognized or marked as OOV	
67	537	272	60	205	61.82%

The errors which lead to incorrect system response can be broadly classified into ASR related and non-ASR related errors.

ERROR CONTRIBUTION FROM DIFFERENT SOURCES



Performance of accent-independent word-based ASR

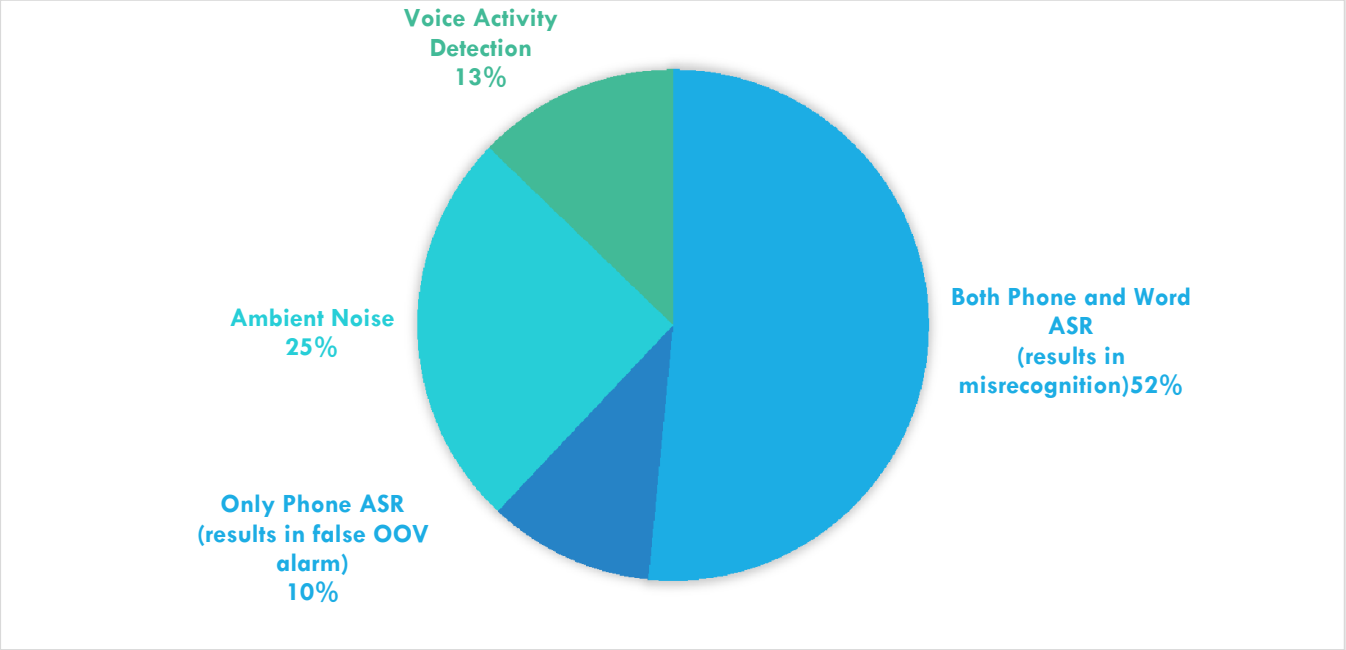
Test Files	Correctly Decoded	Incorrectly Decoded	Accuracy of Word ASR
379	320	59	84.43%

FIELD ACCURACY OF DIALOG SYSTEM WITH ACCENT DEPENDENT ASR

In case of dialog system with accent dependent ASRs, the errors due to non-ASR issues (voice activity detection and background noise) remain the same but errors due to speech recognition system increase significantly and we get an overall drop in the accuracy of the complete system.

No. of Speakers	Total Test Files	Correct System Response		Incorrect System Response	Overall System Accuracy
		In-vocabulary word correctly decoded	OOV or Multiple words correctly identified	In-vocabulary words misrecognized or marked as OOV	
67	537	219	60	258	51.95%

ERROR CONTRIBUTION FROM DIFFERENT SOURCES



Performance of accent-independent word-based ASR

Test Files	Correctly Decoded	Incorrectly Decoded	Accuracy of Word ASR
379	246	133	64.91%

CONCLUSION

In field, accent-independent ASRs outperform the accent-dependent ASRs.

FUTURE WORK

In order to minimize the gap between ASR results in lab and in field, We will improve the accuracy of:

- Baseline ASR systems
- Out of vocabulary detector
- Accent identification system
- Voice activity detector