Emotion Sensitive Speech Control and Noise Reduction in Minimally Invasive Surgery

Lindsey Dean

Project Statement

- Propose to integrate speech recognition software into Da Vinci robot for control of novel functions

Problem

- Voice control has been part of other systems such as AESOP but has failed due to:
  - Long reaction time
  - Limited reliability
  - User dependent interface

Areas of Research

- Emotion Sensitive Speech Control
  - **Emotion Sensitive Speech Control for Human-Robot Interaction in Minimal Invasive Surgery**
- Noise Reduction Strategies
  - **Speech Control in Surgery: A field Analysis and Strategies**
- Both papers working on SIMIS (Speech in Minimally Invasive Surgery) Database – research tool

Emotion Sensitive Speech Control

- social competence can be integrated into speech control through emotional recognition
- allow interface to recognize emotion, can be utilized in cases where surgeon sounds angry or confused to initialize the computer to ask for feedback confirmation

Noise Reduction Strategies

- Optimize noise reduction by qualifying “silence” found in live OR
- Use feature enhancement algorithms to improve accuracy of recognition in noisy environments OR environment

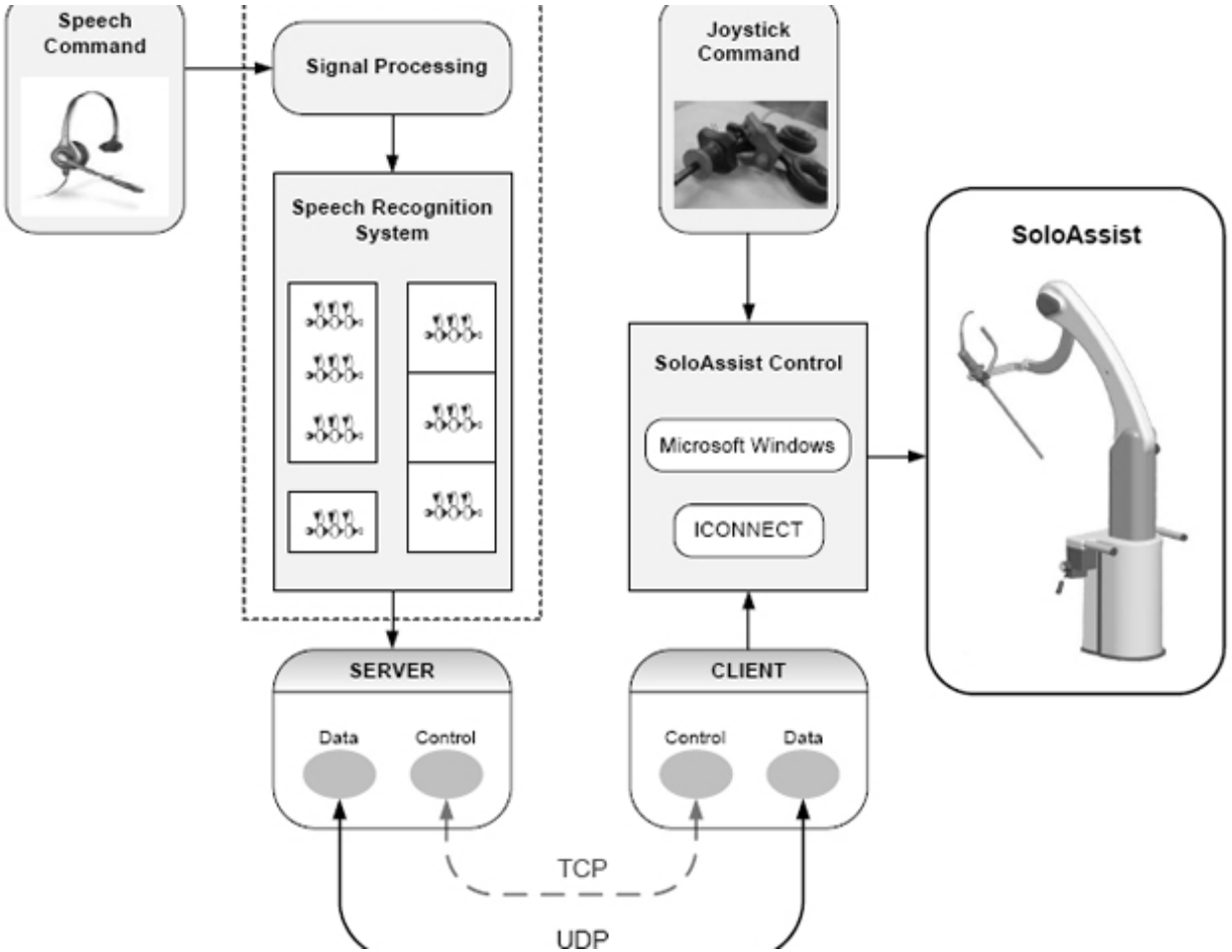
SoloAssist

SIMIS Database

EXPERIMENTAL SET-UP

SoloAssist – AktorMed Germany

Speech Control Interface for SoloAssisst



SIMIS Database

- Surgery in Minimally Invasive Surgery
- 20 recording of live minimally invasive surgeries
  - Segmentation of speech
  - Emotion recognition: labeling of emotion classes of speech
  - Noise reduction: distinguish “noise” sounds from words spoken to robot and label

KEY RESULTS

Table IV – Distribution of Speech Turns Among Emotion By Time and Turn Number Within SIMIS Database

	[m:s]	speech [m:s]	#	neutral [m:s]	#	happy [m:s]	#	angry [m:s]	#	impatient [m:s]	#	confused [m:s]	#
Gall	36:49	6:05	190	2:30	69	1:13	48	0:58	26	0:54	31	0:30	16
	76:14	8:13	308	4:29	151	1:01	56	1:06	34	1:23	57	0:14	19
	34:24	4:45	159	3:18	109	0:24	18	0:30	15	0:09	5	0:24	12
	36:36	8:41	257	6:11	174	1:47	49	0:21	7	0:38	18	0:15	10
Fundoplicatio	54:33	15:05	456	8:26	248	1:01	41	1:57	51	2:30	75	1:11	41
	76:25	16:44	523	10:31	331	1:22	57	1:23	37	2:05	54	1:23	44
Sigma	80:08	14:03	201	7:35	97	1:19	21	1:08	19	1:20	19	2:41	45
Wedge	53:59	12:01	340	7:04	189	1:14	43	0:34	22	1:57	53	1:00	33
	53:51	13:22	295	9:04	204	0:47	22	0:57	15	1:35	31	0:59	23
Stomach	71:01	15:18	306	6:25	121	2:15	48	2:05	39	2:59	62	1:34	35
total	574:00	114:17	**3035**	65:33	**1509**	15:45	**403**	10:09	**265**	15:30	**405**	10:11	**278**

Emotion Sensitive Speech Recognition: Key Results

- Labeled 3035 words of spontaneous real-life speech from OR
  - Only 53% of surgeon-robot interactions were labeled neutral

Emotion Sensitive Speech Recognition: Key Results continued

- Can constrain emotional mapping to two dimensions: neutral and positive vs. negative
  - Discriminate only where a feedback dialogue needs to be initialized

%	Ave	Std dev	Max
RR	75.5	7.7	92.5
CL	71.4	10.6	92.3
F1	73.3	9.1	92.4

Modeling Silence in the Operating Room

- Standard background noise
- Instrument click noise
- Background talk
- Stressed breath or cough

	Turns	Turns/OP	Distribution (%)	Time [m:s]
Std.background	19855	**993**	57.9	583:07
Instr.clock	7839	**392**	22.9	230:13
Bkgrd.talk	3015	**151**	8.8	88:31
Str.breath	3575	**179**	10.4	105:02
Total	34284	**1715**	100	1006:53

Noise Reduction Strategies: Key Results

Table 3: Accuracies for Different Noise Reduction Methods and Noise Types

	MFC	PLP	NT	CMS	HEQ	SDM
clean	**98.53**	98.16	97.06	87.50	97.43	92.96
High SNR	92.59	92.96	97.06	81.99	**95.96**	92.52
Med SNR	90.49	90.49	95.22	79.78	**95.22**	90.15
Low SNR	89.34	89.63	94.12	79.04	**93.75**	88.56
Std.bkgrd	91.65	92.65	95.59	86.4	**97.06**	**92.11**
instr,.click	89.34	89.63	95.96	81.62	**94.12**	**92.22**
Bkgrd.talk	89.71	89.71	94.85	80.51	**94.41**	88.42
Str.breath	79.41	79.62	90.07	77.57	**90.81**	**85.84**
mean	90.13	90.36	94.99	81.80	**94.84**	**90.35**
Weighted mean	86.67	90.34	95.03	83.87	**95.50**	**91.16**

Relevance and Further Work

ASSESSMENT

Relevance to Class Project

- Not planning to work directly with speech recognition engine however important to understand how they work and features of voice
- Measuring silence in the operating room directly
- Emotion sensitive speech recognition beyond scope
  - However feedback from robot based on voice in important
- Vocabulary developed for speech interface
- Security feedback based on AI related to surgeons emotions, vocal feedback
  - May not need to be based on acoustic features but rather out of the ordinary decisions
- SIMIS database potential research tool

Possible Steps to Further Work

- Noise reduction modeled noise with Gaussian would be interesting to use other method to model noise
- Expand SIMIS database
- Design program that uses precedent emotion to predict next since emotions do not change frequently in real life (reduce search size)

Bibliography

- Schuller, Bjorn, Gerhard Rigoll, Salman Can, and Hubertus Feussner. "Emotion Sensitive Speech Control for Human-Robot Interaction in Minimal Invasive Surgery." *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*(2008): 453-58. Print.
- Schuller, Bjorn, Salman Can, Hubertus Feussner, Martin Wollmer, Dejan Arisc, and Benedikt Hornler. "SPEECH CONTROL IN SURGERY: A FIELD ANALYSIS AND STRATEGIES." *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on* (2009): 1214-217. Print.