Rakuten rapidapi – for speech to text api (googlecloudspeech for performance or googletranslate for multi -language capabilities, install both and have user specify which one to use?)

Semantic memory computer model - <https://www.sciencedirect.com/science/article/pii/S0022537169800691?via%3Dihub>

LAIR Research - <http://lair.cse.msu.edu/pub.html>

<https://journals.lww.com/surgical-laparoscopy/Fulltext/2000/06000/A_Comparative_Experimental_Study_Evaluating_the.13.aspx>

Pre-programmed movements are faster than intuitive

Keep log of requested tasks so they can be easily called upon in future

Store tasks by commonality for faster data retrieval

Verbal commands are processed faster than visual commands for both linear and complex movements when using preprogrammed tasks

Verbal commands were executed much slower when the intuitive tasks were introduced

Decreasing required precision of the movements exponentially decreases time required

Verbal commands better for user due to excessive and strange contortions that are needed to get a point across visually

<https://m-cacm.acm.org/magazines/2013/5/163757-collaboration-with-a-robotic-scrub-nurse/fulltext?mobile=true>

verbal command and execution process:

<https://deliveryimages.acm.org/10.1145/2450000/2447993/figs/f2.jpg>

Depth-segmentation algorithm:

Jacob, M.G., Li, Y., Akingba, G., and Wachs, J.P. Gestonurse: A robotic surgical nurse for handling surgical instruments in the operating room. *Journal of Robotic Surgery 6*, 1, (Mar. 2012), 53–63.

Plan for types of objects, ie. Sharp objects should be handled different than others to avoid harm to surroundings

Involves method of holding the object

Planning path of least resistance to get object to its destination including predictions for where humans/other robots may move to during the time the task is executed

Maybe work on different level than human operators so that the collision possible area is linear in the z direction

Object recognition easiest with scannable code, otherwise deep inventory of different angled pictures of that object

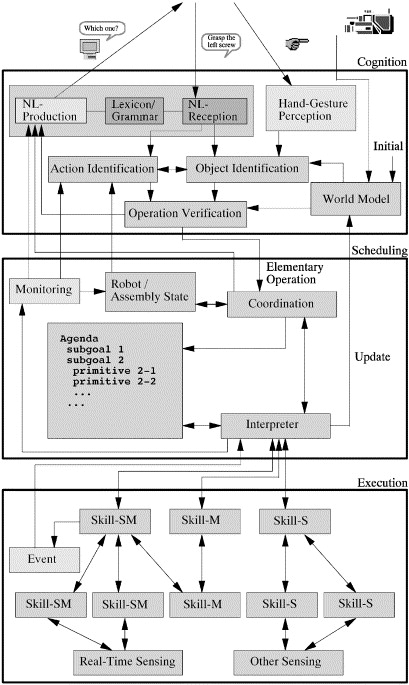
Preform tests for which objects get confused with each other and provide additional information where needed

Recognition should be independent of color due to lighting/contamination

<https://www.sciencedirect.com/science/article/pii/S092188909900041X>

sfb 360 project cited, *check out later*

software set up:



Speech + hand gestures seem to be the most fluid combination of verbal and visual communication

Preprogramming verbs and objects separately and having the software look for these preprogrammed items with the input speech

Build operation with ‘puzzle pieces’

Types of operations:

Preprogrammed

Built

Exceptions (called by another type when some error is thrown)

Target a destination before getting item for easy path processing

Communication between robots to better plan for movements (potentially collaboration such as handing an item to another bot)

Create a database of all objects/obstacles/operators in the immediate (tangible) area

Schedule processes for complex tasks that have been seen before

Maybe user input “New Task Log” to begin recording processes to a script

<https://ieeexplore.ieee.org/abstract/document/7831798>

From command create temp object and add adjectives to compare with database of pre-identified proximal objects

Record nouns for objects, adjectives to identify object, verb to describe task, adverb for things like frequency or repeated tasks, directional terms for location of unknown objects

Treat verbs as functions with different parameters based on the type of action

Parameters being the noun, adjectives, adverbs, directions with only the noun being a non-optional parameter

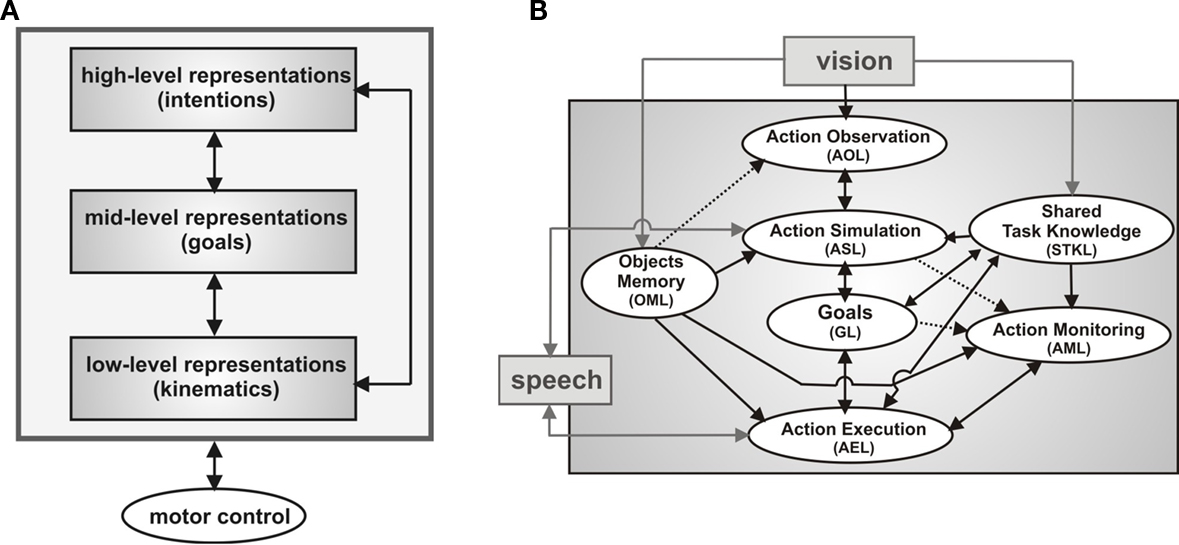
<https://www.frontiersin.org/articles/10.3389/fnbot.2010.00005/full>

this paper has details on deep learning under section: more details

question of whether or not to have the robot speak back (Microsoft Speech SDK 5.1)

responses for commands that throw errors such as moving an object to a destination when the object is already there

study of action-related language (gestures accompanying speech) Glenbach, A. M., and Kaschak, M. P. (2002). Grounding language in action. *Psychon. Bull. Rev.*9, 558–565.



Add a smooth-jerk reaction (make robot take up as little space as possible in the x and y) for scenarios where the coworker makes an unpredicted move

If robot is unsure of an object, point to it and have the human confirm or deny

Predictive action takes into account previous memory of objects’ location/use as well as current visual cues from human – stkl (using 2 connected dnfs)

Create possible goals (by interpreting patterns of the user’s movements) within the next step to speed processing

Example use: prepare next part to be transferred to user before user asks

Correct user if the inputs are illogical; if the user asks for a tool they already have, point to it

Must have an override such as saying “I know” or asking again

Decision making can come either from time it takes to come up with an idea or from weighting different ideas (one works faster but the other is more often correct)

<https://www.aclweb.org/anthology/P17-1150>

objects can have sub-objects in or on them

robot can respond to yes or no questions with a nod of its hand

robot learning is a function of storing new (temp) data into its memory

AI LEARNING - <https://medium.com/technology-invention-and-more/how-to-build-a-simple-neural-network-in-9-lines-of-python-code-cc8f23647ca1>

deep learning memory organized like a graph (process data through graph traversal methods)

think of functions like neurons

for learning give the robot sample data (inputs and outputs) and let it build the equation itself on how the inputs and outputs are related, it tries randomly and adjusts its attempts based on the error of its output vs the correct output

in context of numbers: weight inputs then normalize sum of weight inputs so the output is between 0 and 1: 1/(1+e-x) where x is the sum

adjust weights using formula error\*input\*sigmoidcurvegradient(output)

gradient is small when robot is confident in its weights and large when it’s not

sigmacurvegradient(output)=out\*(1-output)

these links are dead find the papers again

<https://apps-webofknowledge-com.colorado.idm.oclc.org/full_record.do?product=UA&search_mode=GeneralSearch&qid=7&SID=8AcgEWxEiT9ZIy99dtz&page=3&doc=30>

<https://apps-webofknowledge-com.colorado.idm.oclc.org/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=8AcgEWxEiT9ZIy99dtz&page=1&doc=1&cacheurlFromRightClick=no>

<https://apps-webofknowledge-com.colorado.idm.oclc.org/full_record.do?product=UA&search_mode=GeneralSearch&qid=1&SID=8AcgEWxEiT9ZIy99dtz&page=1&doc=2&cacheurlFromRightClick=no>

<https://apps-webofknowledge-com.colorado.idm.oclc.org/full_record.do?product=UA&search_mode=GeneralSearch&qid=7&SID=8AcgEWxEiT9ZIy99dtz&page=1&doc=2&cacheurlFromRightClick=no>

<https://apps-webofknowledge-com.colorado.idm.oclc.org/full_record.do?product=UA&search_mode=GeneralSearch&qid=7&SID=8AcgEWxEiT9ZIy99dtz&page=1&doc=4&cacheurlFromRightClick=no>

<https://apps-webofknowledge-com.colorado.idm.oclc.org/full_record.do?product=UA&search_mode=GeneralSearch&qid=7&SID=8AcgEWxEiT9ZIy99dtz&page=1&doc=5&cacheurlFromRightClick=no>

What we have

Brain structure

Understanding through parts of speech

Speech to text api

What we need

Specific functionalities

Figuring parts of speech from input text - <https://www.aclweb.org/anthology/W19-1603> ?