Investigation of different speech type and emotions for detecting depression using different classifier

by Haihua Jiang, Bin Hu

Review by Farhad Javanmardi

Introduction

- Depression is one the most common mental disorders. It presents persistent feelings of sadness, negativity and difficulty coping with everyday responsibilities. In the worst case, it causes siucide.
- Depression produces both cognitive and physiological impairments.
- Small physiological and cognitive changes will influence:
  - the process of speech production
  - the acoustic quality of the speech

Speech Production

- The process of speech production contains: 1) cognitive planning and 2) complex motoric muscular actions.
- Speech production includes the the formation of the utterance using phonetic and prosodic information.
- This information store in working memory.
- Motor actions part viewed as source-filter operation.
- The major muscle groups used in speech production process are: 1) the respiratory 2) laryngeal and 3) articulatory muscles.

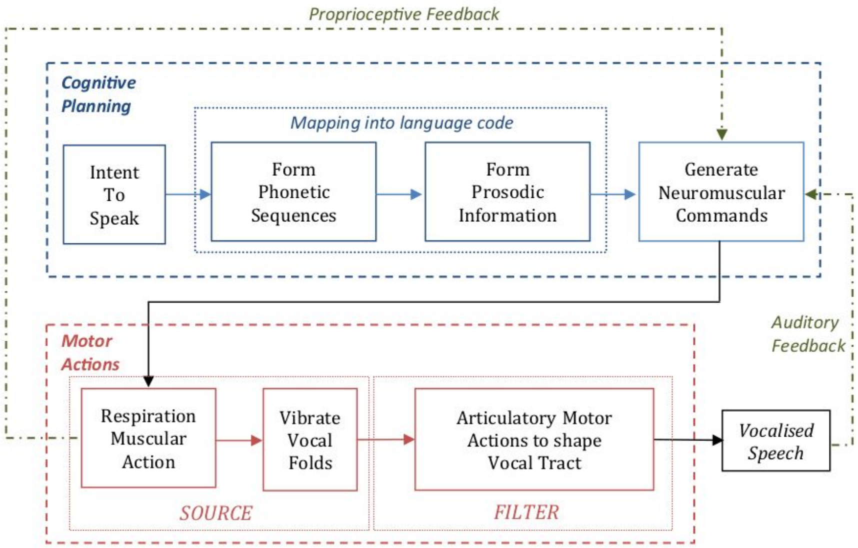

Fig. 1. Schematic diagram of speech production, adapted from Krajewski et al. (2012) and O'Shaughnessy (1999).

Image source Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication. 2015 Jul 1;71:10-49.

Cognitive and physiological impairments

- Cognitive impairments affect the working memory which results in phonation and articulation errors

- Working memory impairments affect speech planning which results in:
    - Pause when talking
    - Difficulty in choosing word

- Changes in somatic nervous system or autonomic nervous system cause:
    - Disturbances in muscle tension and respiratory rate

Cognitive and physiological impairments

- Changes in muscle tension will:
    - Influence the vocal fold behaviour
    - Change vocal tract dynamics
    - Constrain articulatory movement
- These changes affect the prosody and quality of the speech
    - a decrease in speech rate, an increase in hesitation
- Prosodic speech abnormalities such as
    - reduced pitch, reduced loudness, slower speaking rate and articulation errors
- Overally, the depressive voice described as:
    - Speaking in low voice, slowly, hesitatingly, monotonously
    - trying several times before bring out a word
    - becoming mute in the middle of a sentence

The goal

- This paper investigates the impact of speech types and emotions in depression classification and provides an effective measure for detecting depression
- Speech types:
  - Interview, picture description, reading
- Speech emotions:
  - Positive, neutral and negative
- Classifiers
  - KNN, GMM and SVM

Speech database

- 85 depressed subjects (53 females and 32 males)
- 85 healthy controls (51 females and 34 males)
- Each subject's speech divided into 29 recordings
  - Interviews (18 questions = 6 positive + 6 neutral + 6 negative)
  - Picture description (4 pictures = 1 positive + 1 neutral + 1 negative + 1 crying woman picture)
  - Reading (short story with 7 recordings = 2 positive + 3 neutral + 2 negative)

Methodology (pre-processing)

- Group the recordings into three categories according to speech types
  - INT(18 speech recording * 170 participants)
  - PIC (4 speech recording * 170 participants)
  - REA (7 speech recording * 170 participants)

- Group the recordings into three categories according to the emotions
  - POS (6 INTs + 2 REAs + 1 PIC)
  - NEU (6 INTs + 3 REAs + 1 PIC)
  - NEG (6 INTs + 2 REAs + 2 PICs)

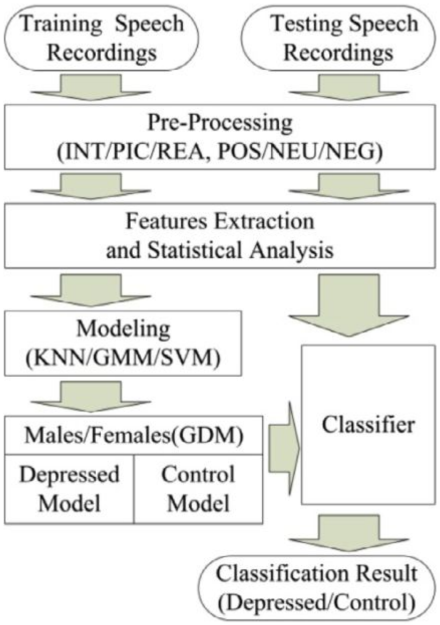

Fig. 1. Block diagram for modeling speech of depressed and control subjects.

Methodology (feature extraction)

- Open-source software openSMILE
- 34 low-level descriptor with 34 corresponding delta coefficients
- 4 pitch-based LLDs and their 4 delta coefficient contours
- 21 statistical functions applied over 68 contours

Table 1
Description of the acoustic features based on 38 LLDs and their first derivate and 21 feature statistics functions.

Descriptors	Functions
PCM loudness	maxPos, minPos, mean,
MFCC[0–14]	stddev, skewness, kurtosis,
Log Mel-frequency band[0–7]	quartile 1/2/3
LSP frequency[0–7]	quartile range (2–1)/(3–2)/(3–1)
F0 envelop	lin.regression coeff.1/2
Voicing probability	lin.regression error Q/A
F0final	percentile 1/99
jitterLocal, jitterDDP	percentile range (99–1)
shimmerLocal	up-level time 75/90

Experiments

- There are two different classification techniques:
    - Gender-dependent modeling (GDM)
    - Gender-independent modeling (GIM)
- In this study, gender-dependent modeling (GDM) employed
    - Males and females were modelled separately
- KNN, GMM and SVM models were used to detect depression for
    - Males (Experiment 1)
    - Females (Experiment 2)
- For experiments 3, a combination of classifiers applied (new methodology)

New methodology (training)

- A weighted decision fusion process used
- In the training part, each of the classifiers C_i (i=1,...,6) were trained using the speech of S_i (i=1,...,6)
- Majority vote used to calculate values of class estimate $y_i(x_k)$

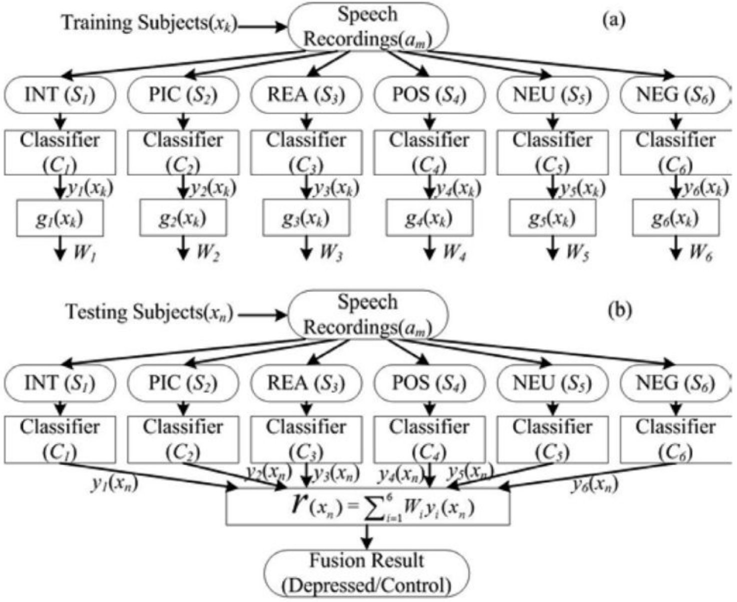

Fig. 2. Overview of the proposed methodology: (a) is the training process and determines the weight values; (b) is the testing process.

New methodology (training)

- The weight coefficient W_i computed as follows:

$$g_i(x_k) = \frac{1}{N}\sum_{k=1}^{N}(y_i(x_k) - f_i(x_k))^2 \qquad (1)$$

$$W_i = \frac{\frac{1}{g_i(x_k)}}{\sum_{i=1}^{6}\left(\frac{1}{g_i(x_k)}\right)} \qquad (2)$$

$$\sum_{i=1}^{6} W_i = 1 \qquad (3)$$

New methodology (testing)

- The class of each subjects x_n in testing set was predicted by each of the six classifiers.
- Each class estimate $y_i(x_n)$ combined with the weight W_i to produce a weighted score parameters $r(x_n)$ as follows:

$$r(x_n) = \sum_{i=1}^{6} W_i y_i(x_n) \qquad (4)$$

- if $r(x_n) > 0$, subject x_n is classified as depressed, otherwise, x_n is classified as healthy control.
- an unweighted decision fusion process (UDD) was also tested for comparison purpose

Results (Experiments 1)

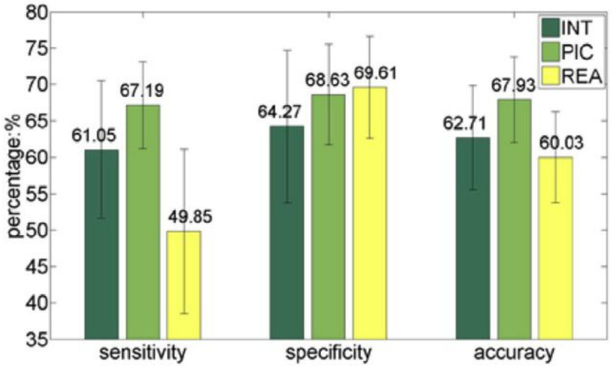

Fig. 3. The mean and st.dev. of sensitivity, specificity, and accuracy of the INTs, PICs, and REAs for males.

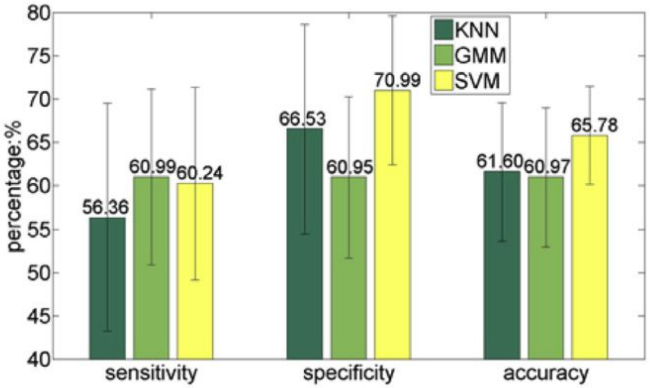

Fig. 5. KNN, GMM, and SVM classification results for males.

- sensitivity, specificity, and accuracy were measured to evaluate the performance of the models.
- Based on result on the left figure, PIC performed better among these speech types.
- From the figure in the right, SVM showed higher value for accuracy.
- After conducting ANOVA and LSD tests, the was $p > 0.05$, meaning that the classification results were similar for speech emotions.

Results (Experiments 2)

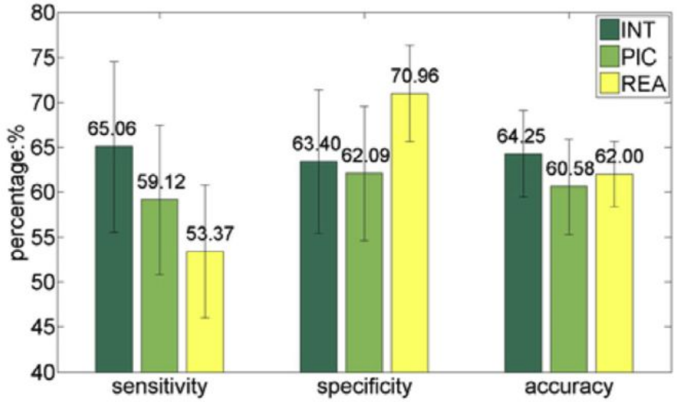

Fig. 6. The mean and st.dev. of sensitivity, specificity, and accuracy of the INTs, PICs, and REAs for females.

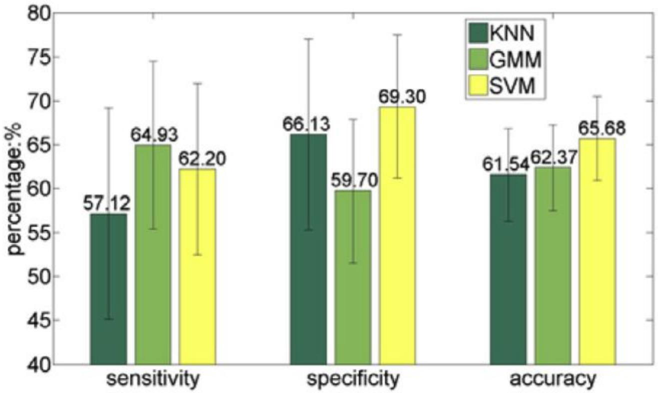

Fig. 8. KNN, GMM, and SVM classification results for females.

- Based on the left figure, INT performed better than PIC for females.
- From the figure in the right, SVM resulted in higher accuracy over KNN and GMM.
- After conducting ANOVA and LSD tests, the was $p > 0.05$, meaning that the classification results for POS, NEU and NEG were similar for females.

Results (Experiments 3)

- Based on the weight values on table 6, for male, PIC has the highest value, meaning that it has the highest correlation with the depression classification.
- For females, INT has the highest value, meaning that the features from INT were more highly correlated with the depression classification.
- From the results in the table 7, it can be seen that the proposed method STEDD (speech types and emotions detection for detection depression) was more effective in detection of depression for both males and females.

Table 6
Weight values for STEDD.

Gender	INT	PIC	REA	POS	NEU	NEG
Male	0.1474	0.1751	0.1273	0.1751	0.1751	0.2001
Female	0.1937	0.1326	0.1292	0.1481	0.2099	0.1866

Table 7
Classification results of UDD and STEDD.

Classifier	Gender	Sensitivity %	Specificity %	Accuracy %
UDD	Male	75.00	79.41	77.27
	Female	73.58	64.71	69.23
STEDD	Male	75.00	85.29	80.30
	Female	77.36	74.51	75.96

References

[1] Jiang, Haihua & Hu, Bin & Liu, Zhenyu & Yan, Lihua & Wang, Tianyang & Liu, Fei & Kang, Huanyu & Li, Xiaoyu. (2017). Investigation of Different Speech Types and Emotions for Detecting Depression Using Different Classifiers. Speech Communication. 90. 10.1016/j.specom.2017.04.001.

[2] Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication. 2015 Jul 1;71:10-49.

[3] Krajewski, Jarek & Schnieder, Sebastian & Sommer, David & Batliner, Anton & Schuller, Björn. (2012). Applying multiple classifiers and non-linear dynamics features for detecting sleepiness from speech. Neurocomput.. 84. 65-75. 10.1016/j.neucom.2011.12.021.

Questions

1. Explain briefly how emotional state of a person suffering from a depressive disorder affects the acoustic qualities of his/her speech?

2. What are the 4 pitch-based low level descriptors in this study and explain them briefly?