Do social anxiety individuals hesitate more? The prosodic profile of hesitation disfluencies in Social Anxiety Disorder individuals

Vered Silber-Varod [1], *Hamutal Kreiner*[2], *Ronen Lovett*[2], *Yossi Levi-Belz*[2], *Noam Amir* [3]

[1] The Research Center for Innovation in Learning Technologies, The Open University of Israel
[2] Linguistic Cognition Lab, Ruppin Academic Center, Israel
[3] Dept. of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University

vereds@openu.ac.il, hamutalk@ruppin.ac.il; ronenlovett@ruppin.ac.il;
Yossil@Ruppin.ac.il, noama@post.tau.ac.il

Abstract

Building on psychologists' observations that individuals with Social Anxiety Disorder (SAD) speak slower and more quietly, this study examines to what extent the characteristics of hesitation disfluencies and silent pauses distinguish between SAD and control participants. Participants responded verbally to six identical questions, and their responses were recorded and analyzed. Our first observation was that SAD sessions last longer. When looking at inter-pausal units, silent pauses, and hesitation disfluencies, we found comparable proportions of hesitation disfluencies in both groups. Critically, however, we found that SAD sessions last longer, due both to more speech and to more silences. A more detailed acoustic analysis examined four types of hesitations with respect to their syntagmatic location, i.e., their location with regard to the speech unit. Results show differences between SAD and control participants in duration, jitter and shimmer. The findings suggest that acoustic analysis of speech disfluencies may serve as an important clinical aid in the diagnosis of SAD.

Index Terms: social anxiety disorder, silences, filled pauses, hesitation disfluencies, acoustic analysis

1. Introduction

Social Anxiety is one of the most common mental disorders in the Western world [1, 2]. It causes individuals psychological distress, and results in significant economic damage due to the lost workdays [3]. A prominent characteristic of Social Anxiety Disorder (SAD) is the difficulty in verbal communication. Individuals with SAD are afraid that others will notice their vocal tremor, that they will say something embarrassing or have a "Black-Out", that they will talk foolishly or will no longer be able to speak [4, 5]. Treatment of SAD is relatively effective, however many individuals with SAD are not diagnosed as such. One of the fundamental problems in diagnosing SAD is that current diagnostic tools rely on self-report, and require the patient to disclose his/her distress. Such self-disclosure is essentially in contrast to the typical "shy" behavior of individuals with SAD. Consequently, many individuals with SAD are not diagnosed [7, 8], and thus do not receive treatment.

Aiming to find different ways to diagnose and evaluate SAD, previous research examined the use of physiological measures. Indeed, several studies show that physiological measures are sensitive not only to short term variation in arousal and emotion, but also to long lasting emotional states. For example, in a study that recorded various physiological measures while participants were presented with provoking stimuli, SAD patients responded with higher levels of skin conductance, heart rate, and muscular tension compared to control group. Moreover, only SAD patients showed increased responses when imagining either idiographic fear or standard social threat scenes. The groups did not differ however when imagining contents which is fearful for all participants [9].

Recent evidence suggests that acoustic parameters are associated with the physiological responses to emotional stress and social anxiety [2]. In particular, several studies [10, 11, 12, 13] found that social anxiety is associated with higher f0 and higher speech rate, as well as a larger range of f0. Voice tremor, which is an acoustic parameter related to instability of f0 (known as Jitter), was reported to be associated to SAD and other emotions [14, 11, 15, 16, 17]. Apart from f0, other studies showed that individuals with SAD use longer silent pauses than control participants [18]. One study [19] found a significant correlation between the existence of social anxiety and the use of various pauses: Participants with SAD used more filled pauses and their silent pauses were longer than in the control group. Hence, a closer examination of the pauses may allow better discrimination between individuals with SAD and control participants.

Previous work on the relationship between disfluencies and speech disorders discussed the influence of typology and grammatical classes on the occurrence of speech disruptions of stuttering and fluent children. Disruptions of speech flow were assumed to be differentiated according to their typology [20], such that some disruptions are common to all speakers and fundamentally reflect linguistic uncertainty and imprecision, while others are intended to improve the message comprehension. The disruptions that were considered typical are hesitations, interjections, revisions, un-finished words, and word, phrase or segment repetition. The results of [20] indicated that the stuttering and fluent children do not differ with regard to the occurrence of typical disfluencies. The same results were reported on cluttering disorder [21], where no statistically significant difference was found between the groups concerning the number of hesitations, unfinished words and segment repetitions [21].

The present study focused on acoustic analysis of hesitation disfluencies (AKA filled pauses) in individuals with SAD. It examined to what extent detailed analysis of different types of hesitation disfluencies may distinguish between individuals with SAD and control participants. Two complementary aspects were examined: 1) temporal features, i.e, amount and duration of occurrences; 2) basic acoustic features based on f0. The

findings may contribute to SAD diagnosis, so that it would not depend solely on self-report, but would be based also on objective physical measures, similar to methods applied in emotion detection [15, 22, 23]. Moreover, future research that will develop computerized algorithms for detection of filled pauses [24, 25] will further contribute to the development of automatic methods for SAD diagnosis and promote its treatment.

2. Method

2.1 Participants

Twenty Hebrew speakers participated in the present study. Based on the Liebowitz Social Anxiety Scale (LSAS) self-report version [26, 27, 28], ten participants that scored above 79 were assigned to the SAD group and ten that scored below 32 were assigned to the control group. Each group included five women and five men, with age range 21-54 (mean 30.9).

2.2 Procedure

All participants were told that the aim of the research was to examine interpersonal speech communication differences. They were further informed that they will be interviewed by the experimenter and that the interview will be recorded. All participants read a detailed description of the study and signed an Informed Consent form. The session began with a structured interview comprising six fixed questions (detailed below). The interviewer was a student in the second year of MA program in clinical psychology. Following the interview, participants filled in a digital LSAS self-report version.

2.3 The interview

The interview began with casual "small talk" questions and the following questions gradually increased the degree of self-disclosure expected (Q1: What do you think about the weather today? Q2: What do you think about reality shows on TV? Q3: Tell me about your hometown and the neighborhood where you were raised. Q4: Tell me about a meaningful person in your life. Q5: What part does/has this person play/played in your life? Q6: Tell me about positive and negative qualities of this person.) Each interview was recorded with Sennheiser MKE 2 microphone digitized with an Icicle 48V external sound card connected to a computer. The microphone was at a fixed distance from the speaker's mouth, and the recording was carried out with a sampling frequency of 48 kHz, 16 bit sample resolution .

2.4 Research design

The main unit of analysis in the current research is the Hesitation Disfluency (HD), which consists of a single (C)V syllable. In the present research, we hand-labeled four different types of HD, according to their syntagmatic relations with the Inter Pausal Unit (IPU):

- HDs that were uttered at the beginning of an IPU, without any silent pause (SIL) between the two, were tagged as *Initial [e]* (I).
- HDs that were uttered at the end of an IPU, without any silent pause between the two, were tagged as *Final [e] (F[e])*.
- HDs that were uttered as an elongated syllable were tagged as *Syllable lengthening (Syll)*.
- HDs that were uttered between silent pauses were tagged as *Filled Pauses (FP)*.

Three out of these four – Initial [e], final [e], and filled pauses – are always pronounced with the vowel [e] (For detailed description see [30]). Only syllable lengthening may occur as a CV syllable with other vowels as well, but mostly with [a] or [e] [31]. These parameters were annotated and segmented manually using the PRAAT textgrid tool, by two annotators, the experimenter and an expert phonetician.

Following the segmentation, the acoustic parameters that were extracted were: Number and durations of each of the four HDs, mean values of F0, intensity, jitter and shimmer.

In addition, the following parameters were summarized:

- Number of speech units, defined as Inter pausal units (IPU) of each participant, and their duration.
- Number and duration of silent pauses (SIL) of each participant. The threshold for minimum SIL was set at 70ms. (Occurring only once in the corpus). All five SILs below 100ms were between FP and IPU or between truncated speech and an IPU; Maximum SIL was 30 seconds, however, all 7 cases above 10 seconds were of a single SAD speaker (#1669) in questions 5, 6, and between FP and IPU.
- Number and duration of giggles of each participant.

It should be mentioned that although almost every answer began and ended with a silent interval, these interval lengths are arbitrary and depend on how long the experimenter waited until he started the recording and stopped it. Therefore, for the purpose of this study, we defined start and end of an answer with a vocal interval (speech, hesitation disfluency or a giggle).

3. Results

When measuring the answer length and the average of each answer length, the results show that the total absolute duration of answers were longer for SAD (1913 seconds) compared to control participants (1029 seconds). Average length of answers by SADs was in one case (question 3) 1.5 times the average length of the control group; and in other cases even more, up to four times in question 2 (Table 1).

Table 1: *Average duration (seconds +standard deviation) of the six answers, for SAD and Control groups.*

Question	Duration of SAD answer in seconds (STD)	Duration of Control answer in seconds (STD)
1	9.447 (7.723)	3.262 (2.877)
2	23.628 (21.453)	6.005 (3.856)
3	29.949 (16.789)	19.001 (8.105)
4	37.397 (9.991)	25.087 (13.589)
5	35.173 (23.748)	18.353 (7.843)
6	55.857 (27.388)	31.227 (14.533)

Table 2: *Total occurrences (%) and total duration (in seconds) and ratios (%)) of Speech (IPU), silences (SIL), four types of hesitation disfluencies, and giggle, in the two examined groups.*

Tag	SAD			Control		
	Occurrences (%)	Duration (s)	Duration (%)	Occurrences (%)	Duration (s)	Duration (%)
IPU	39%	948.378	50%	39%	563.703	55%
SIL	36%	679.964	36%	34%	311.412	30%
FP	6%	98.232	5%	6%	51.672	5%
Syll	6%	65.172	3%	6%	44.149	4%
Final [e]	6%	48.679	3%	6%	28.203	3%
Initial [e]	4%	43.270	2%	4%	20.019	2%
Giggle	2%	29.009	2%	1%	10.018	1%
Total		1913			1029	

Given this data, we probed further to examine the cause behind this difference. Is it due to longer speech stretches, hesitations or silent pauses? Our hypothesis was that SAD participants were more hesitant and more cautious than control participants. Hence, we expected to find more silent pauses and more hesitations in their answers.

In the sequel, we first present the average distribution of IPU, silent pauses and four HDs in each group, and the average duration of these variables. We then present the acoustic features that were extracted. For each acoustic variable, we calculated the average per participant and used *T-Test* for independent groups to examine the statistical significance of differences between the groups (significant results of t-tests are marked in the figures with an asterisk [*]).

3.1 Distribution of IPUs, SILs, and HDs

Table 2 presents absolute and relative (in percentages) durations of all annotated labels. In terms of occurrences, both groups have a very similar percentage of IPUs, silence intervals and hesitation tokens. Interestingly, speakers in the SAD group had longer overall absolute durations for both IPU and SIL, compared to the control group. However, when looking at the *relative* durations of each annotated label there are some minor differences. The control group has relatively longer IPUs (55% versus 50%) while the SAD group has relatively longer silences (36% versus 30%). These relative differences were not found to be statistically significant, Giggle ratios were also different, but with minor presence, and will not be discussed further.
The next section addresses acoustic measurements of the four types of hesitation disfluencies.

3.2 Prosodic profile of hesitation types

In this section the results will be described in terms of average values and t-test results. Figure 1 presents the average durations of the four HDs in the two groups over all six interview answers. Average was calculated first per participant, and then for each group. Three types of HDs (F[e], I and Syll) have similar average duration (the difference was found insignificant at a significance level of 0.05); the difference was found significant for FP ($p = 0.026$).

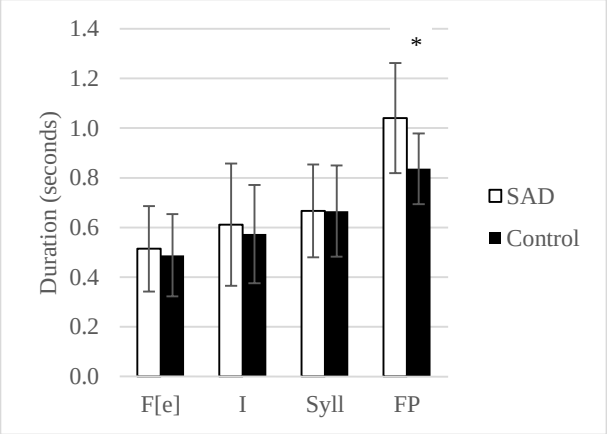

Figure 1: *Mean and 95% confidence interval of duration (s) of hesitation disfluencies: Final [e] (F[e]), Initial [e] (I), Syllable lengthening (Syll), and Filled pauses (FP) (* indicates significant difference), in the two examined groups*.

The mean intensity measurements showed mixed directions. Two types, Initial [e] and syllable lengthening, were pronounced at lower intensity by the control group, while Final [e] and FP were pronounced with lower intensity by the SAD group. With respect to the HDs types, Syll had the lowest mean intensity values for Controls, while Syll and FP were pronounced with lowest intensity by SADs. However, in t-tests these tendencies did not show any significant difference between the two groups.

The mean f0 measurements showed very similar f0 values between the two groups (Figure 2). Females in the SAD group pronounced Initial [e] with higher f0 than the other three types. This can be related to the initial f0 bootstrapping that the speaker does at the beginning of utterances.

The mean jitter measurements show that the SAD group consistently pronounced HDs with more jitter than the control group (Figure 3). The largest difference was found for Syll, which was also the only HD for which the difference was significant. Jitter in Syll is the highest for SADs, while jitter of Initial [e] is highest for Controls. The mean shimmer measurements show similar directions as for jitter: SAD group pronounced HDs with consistently more shimmer than the control group (Figure 4). The largest difference was found for Syll, as in jitter. This difference was significant, as was also the difference for FP.

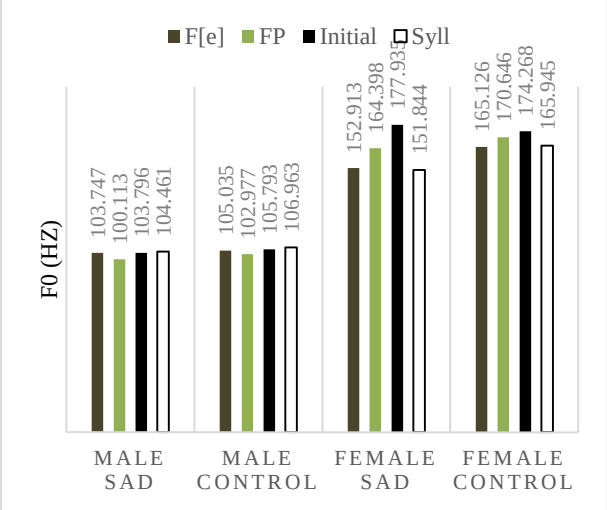

Figure 2: *Mean f0 (Hz) of hesitation disfluencies: Bars represent final [e] (F[e]), Initial [e] (I), Syllable lengthening (Syll), and Filled pauses (FP), in the two examined groups*.

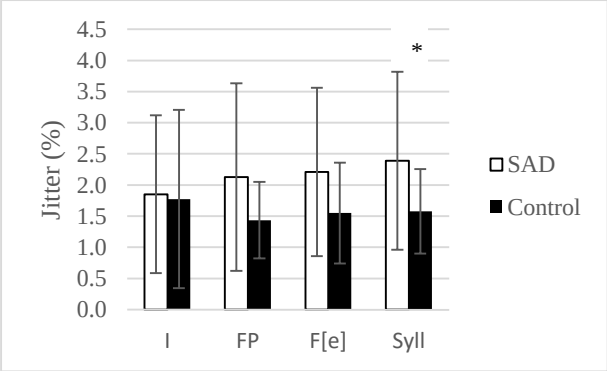

Figure 3: *Mean and 95% confidence interval of jitter (%) of: Final [e] (F[e]), Initial [e] (I), Syllable lengthening (Syll) (* indicates significant difference), and Filled pauses (FP), in the two groups.*

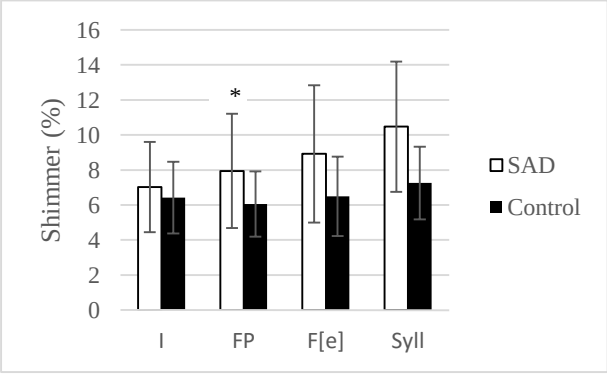

Figure 4: *Mean and 95% confidence interval shimmer (%) of: Final [e] (F[e]), Initial [e] (I), Syllable lengthening (Syll)) (* indicates significant difference), and Filled pauses (FP), in the two groups.*

4. Discussion

In this study we examined to what extent and in what ways the acoustic properties of hesitation disfluencies could be used as cues for diagnosis of SAD. For this purpose, we defined four types of hesitations with respect to their syntagmatic location, i.e., their location with regard to the speech unit (IPU). First, we found comparable hesitation ratios in both groups. This finding is in contrast to [19], who found that SAD subjects use more filled pauses than Controls. These inconsistent results may be due to cross-language differences or to the different nature of the speech interaction (public vs. personal interaction), further research may shed light on the effect of such factors. Critically, however, we found that SAD sessions last longer, due both to more speech and to more silences. This finding is consistent to previous findings reported in [18]), though not in identical proportions. These findings are in agreement with previous research that reported that SAD subject tend to use more silent pauses compared to average population [18], and to show increased responses [9], especially when they are requested to give personal information, as was the case in the current study. Finally, acoustic analysis of five prosodic variables: Duration, f0, intensity, jitter, and shimmer, showed that the two groups differ in the way they pronounce HDs only to a limited extent. Overall, SAD individuals were found to have longer filled pauses; higher jitter and shimmer in syllable lengthening and higher shimmer in filled pauses. Intensity was not found a significant parameter to discriminate between the two groups.

The main differences that we found between the four types of HD were: Filled pauses were the longest, while Final [e]s the shortest. Jitter and shimmer were similar in all four types of HD, but only in the Control group; SAD group produced the syllable lengthening type with the highest rate of jitter and shimmer.

The contribution of the current study is twofold: First, by a new typology of hesitation disfluencies, based on syntagmatic position. This annotation is straightforward and may be of use for automatic detectors; Second, the findings suggest that whereas quantitatively similar proportions of HDs are produced in both groups, qualitatively the two groups differ in some aspects of the HDs. Thus, acoustic analysis of speech disfluencies may serve as important tool for objective diagnosis of SAD, and the results can be used to develop potential SAD classifier.

In future research we intend to concentrate on the acoustic measurements of the speech units (IPUs), as well as transcribe the interviews in order to further investigate the syntagmatic aspect of HDs with respect to its interface with the linguistic level. This may allow us to examine in what ways SADs differ from Controls in their use of language.

5. References

[1] R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, & E. E. Walters, "Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication," *Archives of General Psychiatry*, vol. 62, no. 6, pp. 593–603, 2005.

[2] J. W. Weeks, C. Y. Lee, A. R. Reilly, A. N. Howell, C. France, J. M. Kowalsky, & A. Bush, "The sound of fear: Assessing vocal fundamental frequency as a physiological indicator of social anxiety disorder," *Journal of Anxiety Disorders*, vol. 26, no. 8, pp. 811–822, 2012.

[3] M. B. Stein, & D. J. Stein, "Social anxiety disorder," *The Lancet*, vol. 371, no. 9618, pp. 1115–1125, 2008.

[4] F. R. Schneier, "Social anxiety disorder," *New England Journal of Medicine*, vol. 355, no. 10, pp. 1029–1036, 2006.

[5] M. B. Stein, J. R. Walker, & D. R. Forde, "Public-speaking fears in a community sample: Prevalence, impact on functioning, and diagnostic classification," *Archives of General Psychiatry*, vol. 53, no. 2, pp. 169–174, 1996.

[6] T. Pierce, "Social anxiety and technology: Face-to-face communication versus technological communication among teens," *Computers in Human Behavior*, vol. 25, no. 6, pp. 1367–1372, 2009.

[7] M. Olfson, M. Guardino, E. Struening, F. R. Schneier, F. Hellman, & D. F. Klein, "Barriers to the treatment of social anxiety," *American Journal of Psychiatry*, vol. 157, no. 4, pp. 521–527, 2000.

[8] P.S. Wang, M. Lane, M. Olfson, H. A. Pincus, K. B. Wells, & R. C. Kessler, "Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication," *Archives of general psychiatry*, vol. 62, no. 6, pp. 629–640, 2005.

[9] L. M. McTeague, P. J. Lang, M. C. Laplante, B. N. Cuthbert, C. C. Strauss, & M. M. Bradley, "Fearful imagery in social phobia: generalization, comorbidity, and physiological reactivity," *Biological Psychiatry*, vol. 65, no. 5, pp. 374–382, 2009.

[10] J. Pittam, & K. R. Scherer, "Vocal expression and communication of emotion," In M. Lewis, J. M. Haviland (Eds.), *Handbook of Emotions* (pp. 185–197). New York, NY, US: Guilford Press, 1993.

[11] S. P. Whiteside, "Simulated emotions: an acoustic study of voice and perturbation measures," In: *Proceedings of ICSLP 1998*, PP. 699–703, 1998.

[12] K. R. Scherer, "Vocal affect expression: a review and a model for future research," *Psychological bulletin*, vol. 99, no. 2, p. 143, 1986.

[13] L. Devillers, & I. Vasilescu, "Prosodic cues for emotion characterization in real-life spoken dialogs," In: *Proceeding of the 8th European Conference on Speech Communication and Technology*, Geneva, Switzerland, 2003.

[14] At. Johnstone, & K. R. Scherer, "The effects of emotions on voice quality," In *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 2029–2032). San Francisco: University of California, Berkeley, 1999.

[15] S. M. Yacoub, S. J. Simske, X. Lin, & J. Burns, "Recognition of emotions in interactive voice response systems," In: *Proceeding of the 8th European Conference on Speech Communication and Technology* (pp. 729–732). Geneva, Switzerland, 2003.

[16] J. P. Cabral, L. C. Oliveira, "EmoVoice: A System to Generate Emotions in Speech," In: *Proceedings of Interspeech – ICSLP*, pp. 1798–1801. Pittsburgh, 2006.

[17] S. Patel, K. R. Scherer, E. Björkner, & J. Sundberg, "Mapping emotions into acoustic space: the role of voice production," *Biological psychology*, vol. 87, no. 1, pp. 93–98, 2011.

[18] P. Laukka, C. Linnman, F. Ahs, A. Pissiota, Ö. Frans, V. Faria, A. Michelgard, L. Appel, M. Fredrikson, & T. Furmark, "In a nervous voice: Acoustic analysis and perception of anxiety in social phobics' speech," *Journal of Nonverbal Behavior*, vol. 32, no. 4, pp. 195–214, 2008.

[19] S. G. Hofmann, A. L. Gerlach, A. Wender, & W. T. Roth, "Speech disturbances and gaze behavior during public speaking in subtypes of social phobia," *Journal of Anxiety Disorders*, vol. 11, no. 6, pp. 573–585, 1997.

[20] F. Juste, & C. R. F. D. Andrade, "Typology of speech disruptions and grammatical classes in stuttering and fluent children," *Pró-Fono Revista de Atualização Científica*, vol. 18, no. 2, pp. 129–140, 2006.

[21] C. M. C. D. Oliveira, A. P. L. Bernardes, G. A. F. Broglio, & S. A. Capellini, "Speech fluency profile in cluttering individuals," *Pró-Fono Revista de Atualização Científica*, vol. 22, no. 4, pp. 445–450, 2010.

[22] D. Rochman, G. M. Diamond, & O. Amir, "Unresolved anger and sadness: Identifying vocal acoustical correlates," *Journal of Counseling Psychology*, vol. 55, no. 4, pp. 505–517, 2008.

[23] N. Amir, H. Mixdorff, O. Amir, D. Rochman, G. M. Diamond, H. R. Pfitzinger, T. Levi-Isserlish, & S. Abramson, "Unresolved anger: Prosodic analysis and classification of speech from a therapeutic setting," Paper presented at the meeting of *Speech Prosody 2010*, Chicago, IL, 2010.

[24] E. Shriberg, R. A. Bates, & A. Stolcke, "A prosody only decision-tree model for disfluency detection," In *Eurospeech*, vol. 97, pp. 2383-2386, 1997.

[25] A. Batliner, A. Kießling, S. Burger, & E. Nöth, "Filled pauses in spontaneous speech," Technical Report 88, Verbmobil Project. also in *ICPhS'95*, 1995.

[26] D. S. Mennin, D. M. Fresco, R. G. Heimberg, F. R. Schneier, S. O. Davies, & M. R. Liebowitz, "Screening for social anxiety disorder in the clinical setting: using the Liebowitz Social Anxiety Scale," *Journal of Anxiety Disorders*, vol. 16, no. 6, pp. 661–673, 2002.

[27] N. K. Rytwinski, D. M. Fresco, R. G. Heimberg, M. E. Coles, M. R. Liebowitz, S. Cissell, M. B. Stein, & S. G. Hofmann, "Screening for social anxiety disorder with the self-report version of the liebowitz social anxiety scale," *Depression and Anxiety*, vol. 26, no. 1, pp. 34–38, 2009.

[28] D. M. Fresco, M. E. Coles, R. G. Heimberg, M. R. Liebowitz, S. Hami, M. B. Stein, & D. Goetz, "The Liebowitz Social Anxiety Scale: A comparison of the psychometric properties of self-report and clinician-administered formats," *Psychological Medicine*, vol. 31, no. 6, pp. 1025–1035, 2001.

[29] P. Boersma, & D. Weenink, Praat: doing phonetics by computer [Computer program]. Version 5.3.66, 2014, http://www.praat.org/

[30] V. Silber-Varod, *The SpeeCHain Perspective: Form and Function of Prosodic Boundary Tones in Spontaneous Spoken Hebrew*, LAP Lambert Academic Publishing, 2013.

[31] V. Silber-Varod, "Phonological aspects of hesitation disfluencies," *Proceedings of Speech Prosody 2010*, Chicago, IL, May 2010.