OBJECTIVE ASSESSMENT OF VOCAL TREMOR

Jacob Peplinski[1], *Visar Berisha*[1,2], *Julie Liss*[2], *Shira Hahn*[3],
Jeremy Shefner[4,5], *Seward Rutkove*[6], *Kristin Qi*[6], *and Kerisa Shelton*[4]

[1] School of Electrical Computer and Energy Engineering, Arizona State University, Tempe, USA
[2] Department of Speech and Hearing Sciences, Arizona State University, Tempe, USA
[3] Aural Analytics, Scottsdale, USA
[4] Department of Neurology, Barrow Neurological Institute, Phoenix, USA
[5] University of Arizona College of Medicine, Phoenix, USA
[6] Department of Neurology, Beth Israel Deaconess Medical Center, Boston, USA

ABSTRACT

Detecting early signs of neurodegeneration is vital for planning treatments for neurological diseases. Speech plays an important role in this context because it has been shown to be a promising early indicator of neurological decline, and because it can be acquired remotely without the need for specialized hardware. Typically, symptoms are characterized by clinicians using subjective and discrete scales. The poor resolution and subjectivity of these scales can make the earliest speech changes hard to detect. In this paper, we propose an algorithm for the objective assessment of vocal tremor, a phenomenon associated with many neurological disorders. The algorithm extracts and aggregates a feature set from the average spectra of the energy and fundamental frequency profiles of a sustained phonation. We show that the resultant low-dimensional feature set reliably classifies healthy controls and patients with amyotrophic lateral sclerosis perceptually rated for tremor by speech language pathologists.

Index Terms— Amyotrophic Lateral Sclerosis (ALS), Speech, Tremor, Dysarthria

1. INTRODUCTION

Early detection of neurological disease onset is vital for measuring the efficacy of drug interventions and slowing progression. However, early detection is difficult because the current gold standard endpoints for most neurodegenerative diseases are functional rating scales - questionnaires used by clinicians to assess a patient's symptoms [1][2]. Though these scales span a comprehensive set of disease-specific symptoms, scores for individual questions are low-resolution and subjective, rendering the earliest signs of neurological decline hard to detect. Objective evaluation of speech is gaining popularity as a means of detecting subtle changes in neurological health [3][4][5][6][7]. Several current studies that detail these changes do so in a clinical setting using high-quality microphones [8] or specially-designed hardware and software systems [6][9][10]. Though reliable, these methods are limited in that they still require patients to visit a clinic to be assessed. To increase the sensitivity of early detection paradigms, patients should be able to perform self-administered evaluations remotely and frequently, without the use of specialized hardware. There is some work that proposes the use of mobile devices such as smartphones for collecting clinically-relevant measures of speech and monitoring disease progression [11][12]. However, further work is required to develop tools that reliably extract additional clinically-relevant measures from speech.

Sustained phonation (i.e. prolonging an "ah" for as long and as steadily as possible) is a common speech elicitation task used to evaluate the health of the phonatory speech subsystem. Increased variance in the fundamental frequency (F0) and energy contours of sustained phonations have long been considered interpretable measures of dysarthria [13]. There are several methods for characterizing the variance in F0 and energy in a sustained phonation. The most common family of measures aims to quantify cycle-by-cycle deviations in phonation energy and F0 due to incomplete closure of the vocal folds. Measures in this family include jitter and shimmer [14][15], as well as several more robust measures [16][17]. Though proven to be useful in differentiation of healthy subjects from Parkinsons patients, these measures do not evaluate vocal tremor, a low-frequency modulation in phonation common in neurological disease. Vocal tremor can manifest in some neurological disorders due to a loss of motor control in the laryngeal muscles, resulting in unintended quasi-sinusoidal modulations in energy and F0 [18] [19]. One of the earliest attempts at objectively measuring vocal tremor can be found as part of the Multi-Dimensional Voice Program [20], followed by the work by Brückl et al. in [21][22]. This family of metrics discard information about the distribution of tremor frequencies and intensities in the pitch and energy contours. This presents an issue because tremor frequency and intensity

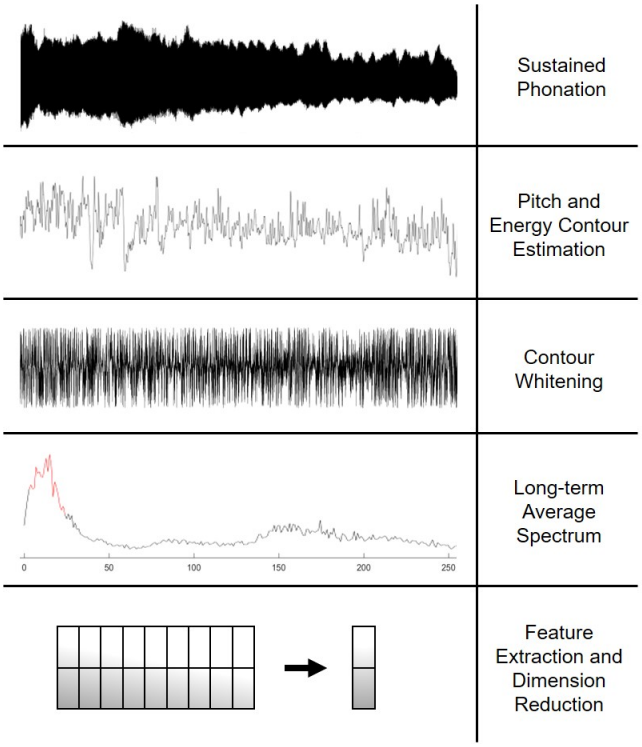

Fig. 1: An overview of the proposed tremor detection algorithm.

can drift in time. To model non-stationary tremors, Pantazis et al. proposed a tremor analysis technique based on the decomposition of sustained phonations into intrinsic mode functions [23]. Each decomposed intrinsic mode function can be analyzed for the presence of tremulous modulations, but no metrics for quantifying the tremor are proposed in that work. Furthermore, it is uncertain which intrinsic mode functions contain the most information on vocal tremor, or how many intrinsic mode functions should be analyzed to fully characterize tremor.

We propose an algorithm that objectively quantifies vocal tremor in sustained phonations. The proposed method estimates fundamental frequency (F0) and energy contours from a sustained phonation and extracts measures of intensity, energy, and entropy to quantify the presence of tremor in energy and pitch. We evaluate this method on a longitudinal speech dataset of 4,834 sustained phonations from 26 healthy English speakers and 65 ALS patients at varying disease stages. We demonstrate the utility of this method by constructing several binary classifiers for separating recordings from healthy controls, recordings from ALS patients rated negative for tremor, and recordings from ALS patients rated positive for tremor using only these proposed measures.

2. ALGORITHM OVERVIEW

Figure 1 provides a high-level overview of the proposed approach. Consider a sustained phonation pre-processed such that silence in the recording and edges of the phonation are removed. We estimate the F0, $p(n)$, and energy, $e(n)$, contour along this phonation and decorrelate both contours via inverse filtering with linear prediction coefficients. We posit that tremor can be characterized by extracting statistics from the average spectra of both contours in the spectral sub-band between 3 Hz and 25 Hz. These bounds are informed by previous literature, both medical [18] and technical [20]. This approach accurately detects the presence of tremor regardless of its non-stationary behavior.

2.1. Pre-processing

Sustained phonations are first processed with a voice activity detector (VAD) to remove all silence segments. We used the VAD method described in [24].

2.2. F0 and Energy Contour Extraction

The VAD-processed signal is decomposed into 10ms analysis windows with a 1ms window overlap. The first and last 5 percent of windows are discarded to avoid amplitude ramping at the beginning and end of the phonation. The signal energy is calculated in each window to form the energy contour of the phonation $e(n)$. Since estimating F0 is a difficult task for speech from clinical populations and the proposed method relies heavily on the shape of the F0 contour, care is taken to ensure that the F0 estimates used herein are reliable. To that end, we extract F0 from each windowed segment of phonation using a modified version of the Praat pitch detection algorithm [25]. The Praat algorithm is modified so that the F0 search range is adjusted depending on the sex of the speaker. This allows us to reduce the frequency search range, which mitigates octave jumping and increases accuracy. The F0 search ranges for males and females were [60-260] Hz and [120-380] Hz respectively. This approach is used to estimate the F0 in successive windows of phonation, forming the F0 contour, $p(n)$.

2.3. Contour Whitening

Forward linear prediction coefficients are estimated by minimizing the expected value of the error between a ground truth sequence and a low-order approximation of the sequence constructed from a linear combination of past samples. Thus, to estimate prediction coefficients a_k for a pitch sequence $p[n]$, we minimize

$$E\left[\left(p[n]-\sum_{k=1}^{q}a_k p[n-k]\right)^2\right], \qquad (1)$$

TABLE I - Tremor Feature Definitions

Feature Name	Mathematical Definition
Dominant Tremor Frequency	$\arg\max X(f_T)$
Max Absolute Tremor Intensity	$\max X(f_T)$
Median Absolute Tremor Intensity	median $X(f_T)$
Mean Absolute Tremor Intensity	mean $X(f_T)$
Max Relative Tremor Intensity	$\max X(f_T)$-median $X(f)$
Median Relative Tremor Intensity	median $X(f_T)$-median $X(f)$
Mean Relative Tremor Intensity	mean $X(f_T)$-median $X(f)$
Tremor Energy	$\|X(f_T)\|_2$
Tremor Entropy	$-\sum X(f_T)\log_2 X(f_T)$

with respect to a_k, where q is the number of prediction coefficients. To obtain the residual contour $p_r(n)$, we inverse filter $p(n)$ with its prediction coefficients

$$p_r[n] = p[n] - \sum_{k=1}^{q} a_k p[n-k]. \quad (2)$$

The inverse filtering detrends each contour, which is speaker-specific; thus, $p_r[n]$ is spectrally-flattened, making it easier to detect the low-frequency tremor-like variations in the signal. The whitening process is repeated for the energy contour to obtain $e_r[n]$, the spectrally-flattened version of the energy contour.

2.4. Spectrum Averaging and Feature Extraction

The signals $e_r[n]$ and $p_r[n]$ are independently decomposed into 1-second windows with an overlap of 100ms. We take the FFT in each window and average across windows to obtain the long-term average spectra $E_r(f)$ and $P_r(f)$ respectively. These average spectra contain information regarding the average intensity and modulation frequencies of $e_r[n]$ and $p_r[n]$. Lastly, we standardize the spectra by z-scoring the values of their spectral bins. To characterize the presence of tremor in the contours, we propose several metrics that capture the prominence of low-frequency variations in the long-term average spectra. Informed by seminal work done on tremor in neurodegenerative disease [18], we restrict the calculation of these metrics to the band range between $f_T \in [3\text{Hz} \ldots 25\text{Hz}]$. These metrics are defined in Table I.

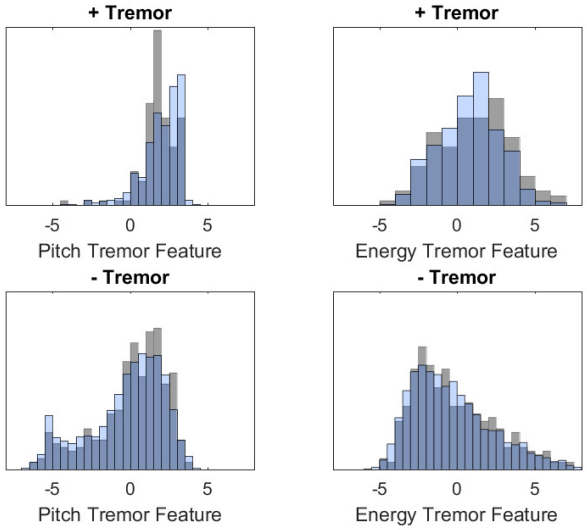

Fig. 2: Feature distributions of clinical and extrapolated tremor ratings. Grey = Clinical, Blue = Extrapolated.

2.5. Dimensionality Reduction

The 9 features detailed in Table I are extracted for $E_r(f)$ and $P_r(f)$. The metrics extracted from the same spectra tend to be highly correlated. Thus, we can reduce the dimensionality of the feature set by collapsing the features into a single feature vector using principal component analysis (PCA). This is done across all sustained phonations in the dataset by first standardizing each feature vector by converting to z-scores then performing PCA. The first PCA dimension is retained as the combination feature vector. The above process is performed for the 9 features on $E_r(f)$ and $P_r(f)$, resulting in two combination features capturing tremor-like characteristics in energy and pitch respectively.

3. EXPERIMENTAL EVALUATION

We evaluate the proposed method on a dataset of sustained phonations recorded by healthy English speakers and ALS patients. The recordings were gathered as part of the ALS at-home study where patients used a speech elicitation tool in a smartphone application [11]. All speech samples were collected with a sampling frequency of 16kHz and 16-bit resolution. Patients were asked to produce sustained phonations of the vowel /a/ once a day (along with other stimuli) for the duration of their participation in the study. The dataset includes 1,650 phonations from 26 healthy patients and 5,017 phonations from 65 ALS patients. Phonations were recorded several times each week over the course of several months. Seven phonations from each ALS patient were assessed for tremor by a speech language pathologist (SLP). Rated phonations were chosen from the beginning, middle, and end of each patient's study participation, so that they are representa-

TABLE II - Performance Metrics for Binary Classification of Tremor Diagnosis Groups

Classes	Sex	Sample Size	Class. Accuracy	ROC Area	FPR
Healthy v. -Tremor	M	2240	58.17	0.580	0.498
Healthy v. +Tremor	M	702	77.07	0.869	0.229
-Tremor v. +Tremor	M	1918	74.14	0.850	0.263
Healthy v. -Tremor	F	2238	57.95	0.601	0.407
Healthy v. +Tremor	F	1384	88.43	0.952	0.118
-Tremor v. +Tremor	F	1326	83.18	0.924	0.172

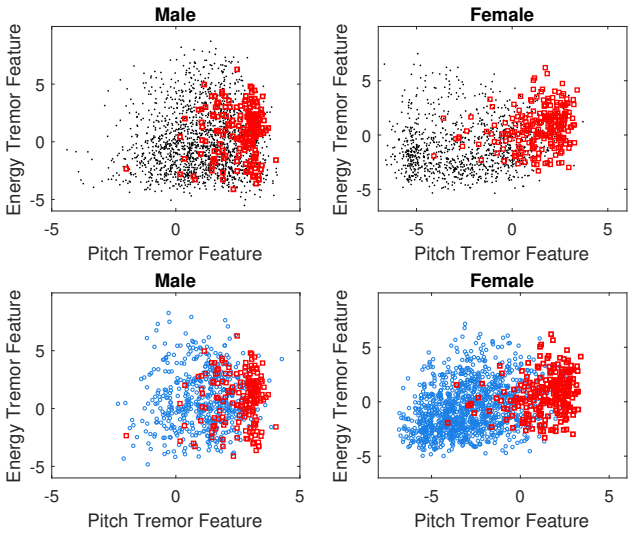

Fig. 3: Plots of tremor features by sex and diagnosis group. Blue = Healthy, Black = -Tremor, Red = +Tremor.

tive of a patient's phonatory ability throughout the study. To increase the statistical power of our analysis, we extrapolate additional tremor ratings using two assumptions. The first assumption is that no ALS patients experienced a decrease in symptom severity over the course of the study. This is reasonable as ALS is a neurodegenerative disease with no expectation that symptoms improve over time. Because tremor is a symptom of neurodegenerative disease, we also assumed that positive ratings for tremor were likely to be followed by subsequent positive ratings for the same participant. Thus, if all seven of a subjects ratings were positive or negative for tremor, we labeled all files for that subject as either positive or negative. The revised set of scores include 423 phonations labeled with tremor and 2,761 phonations labeled without tremor. We confirm that the extrapolated scores represent the SLP-rated scores by plotting histograms of feature values for SLP-rated scores and extrapolated scores in Figure 2.

3.1. Classifying Clinical Ratings using Tremor Features

To demonstrate the utility of the proposed method, we use a collection of binary classification tasks for discriminating sustained phonations labeled by diagnosis group (e.g. healthy control, ALS without tremor, ALS with tremor). Each classification task is run twice: once for each sex. This is because the normative distributions for the tremor scores are not equivalent for males and females. Previous research has also identified significant sex differences in the analysis of F0 and energy contours [26]. We therefore construct six binary logistic regression classifiers which are trained and validated using 10-fold stratified cross-validation. The classification tasks test the discriminative power of the proposed features in classifying all six combinations of two classes (across diagnosis groups and sexes). Because the class sizes are unbalanced, we oversample the minority class using the Synthetic Minority Over-Sampling Technique (SMOTE) proposed in [27]. The SMOTE procedure is only performed on the training set to avoid information leakage between the test and training sets. The results of the classification experiments are presented in Table II. Results of each classification task show that the combination tremor measures possess noticeable discriminative power for separating sustained phonations rated perceptually for vocal tremor. This is true for both males and females, though classification results are much better for women. There is evidence that suggests that measures of deviation in phonatory control may be a better indicator of decline in women compared to men [26]. For both sexes however, the algorithm is incapable of reliably discriminating between healthy controls and ALS patients with negative perceptual ratings for vocal tremor, with predictive power just above that of random guessing. This result is expected due to the fact this method is only intended to measure the presence of tremor. A visualization of tremor feature values for each diagnosis group is shown in Figure 3.

4. CONCLUSION

Current objective measures of tremor do not fully characterize the distribution of tremor frequencies and intensities throughout a sustained phonation. In this paper, we propose a comprehensive approach to characterizing tremor that extracts a variety of statistics from the average spectra of the decorrelated F0 and energy contours of a sustained phonation. We have shown that this approach possesses discriminative power for identifying sustained phonations with perceptual vocal tremor. In conjunction with other objective measures of speech, the proposed measures may be useful in identifying subtle changes in speech, supporting the early detection and possible diagnosis of neurological disease.

5. REFERENCES

[1] J.M. Cedarbaum, N. Stambler, E. Malta, C. Fuller, D. Hilt, B. Thurmond, and A. Nakanishi, "The alsfrs-r: a revised als functional rating scale that incorporates assessments of respiratory function. bdnf als study group (phase iii).," *Journal of the neurological sciences*, vol. 169, no. 1-2, pp. 13, 1999.

[2] C.G. Goetz, B.C. Tilley, S.R. Shaftman, G.T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M.B. Stern, R. Dodel, et al., "Movement disorder society-sponsored revision of the unified parkinson's disease rating scale (mds-updrs): scale presentation and clinimetric testing results," *Movement disorders: official journal of the Movement Disorder Society*, vol. 23, no. 15, pp. 2129–2170, 2008.

[3] V. Berisha, S. Wang, A. LaCross, J. Liss, and P. Garcia-Filion, "Longitudinal changes in linguistic complexity among professional football players," *Brain and language*, vol. 169, pp. 57–63, 2017.

[4] Y. Jiao, V. Berisha, J. Liss, S.C. Hsu, E. Levy, and M. McAuliffe, "Articulation entropy: An unsupervised measure of articulatory precision," *IEEE Signal Processing Letters*, vol. 24, no. 4, pp. 485–489, 2017.

[5] J.R. Green, Y. Yunusova, M.S. Kuruvilla, J. Wang, G.L. Pattee, L. Synhorst, L. Zinman, and J.D. Berry, "Bulbar and speech motor assessment in als: Challenges and future directions," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 14, no. 7-8, pp. 494–500, 2013.

[6] K.M. Allison, T.F. Yunusova, Y.and Campbell, J. Wang, J.D. Berry, and J.R. Green, "The diagnostic utility of patient-report and speech-language pathologists ratings for detecting the early onset of bulbar symptoms due to als," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, vol. 18, no. 5-6, pp. 358–366, 2017.

[7] S. Arora, V. Venkataraman, A. Zhan, S. Donohue, K.M. Biglan, E.R. Dorsey, and M.A. Little, "Detecting and monitoring the symptoms of parkinson's disease using smartphones: a pilot study," *Parkinsonism & related disorders*, vol. 21, no. 6, pp. 650–653, 2015.

[8] J.A. Whitfield and A.M. Goberman, "Articulatory–acoustic vowel space: Application to clear speech in individuals with parkinson's disease," *Journal of communication disorders*, vol. 51, pp. 19–28, 2014.

[9] Y. Yunusova, J.R. Green, J. Wang, G. Pattee, and L. Zinman, "A protocol for comprehensive assessment of bulbar dysfunction in amyotrophic lateral sclerosis (als)," *Journal of visualized experiments: JoVE*, , no. 48, 2011.

[10] P. Rong, Y. Yunusova, J. Wang, L. Zinman, G.L. Pattee, J.D. Berry, B. Perry, and J.R. Green, "Predicting speech intelligibility decline in amyotrophic lateral sclerosis based on the deterioration of individual speech subsystems," *PloS one*, vol. 11, no. 5, pp. e0154971, 2016.

[11] S. Rutkove, K. Qi, K. Shelton, J. Liss, V. Berisha, and J. Shefner, "Als longitudinal studies with frequent data collection at home: study design and baseline data," *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, in press.

[12] A. Tsanas, M.A. Little, P.E. McSharry, and L.O. Ramig, "Accurate telemonitoring of parkinson's disease progression by noninvasive speech tests," *IEEE transactions on Biomedical Engineering*, vol. 57, no. 4, pp. 884–893, 2010.

[13] P. Zwirner, T. Murry, and G.E. Woodson, "Phonatory function of neurologically impaired patients," *Journal of communication disorders*, vol. 24, no. 4, pp. 287–300, 1991.

[14] P. Lieberman, "Some acoustic measures of the fundamental periodicity of normal and pathologic larynges," *The Journal of the Acoustical Society of America*, vol. 35, no. 3, pp. 344–353, 1963.

[15] Y. Horii, "Vocal shimmer in sustained phonation," *Journal of Speech, Language, and Hearing Research*, vol. 23, no. 1, pp. 202–209, 1980.

[16] M.A. Little, P.E. McSharry, S.J. Roberts, D.A.E. Costello, and I.M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomedical engineering online*, vol. 6, no. 1, pp. 23, 2007.

[17] A. Tsanas, M.A. Little, P.E. McSharry, J. Spielman, and L.O. Ramig, "Novel speech signal processing algorithms for high-accuracy classification of parkinson's disease," *IEEE transactions on biomedical engineering*, vol. 59, no. 5, pp. 1264–1271, 2012.

[18] A.E. Aronson, W.S. Winholtz, L.O. Ramig, and S.R. Silber, "Rapid voice tremor, or flutter, in amyotrophic lateral sclerosis," *Annals of Otology, Rhinology & Laryngology*, vol. 101, no. 6, pp. 511–518, 1992.

[19] E.A. Strand, E.H. Buder, K.M. Yorkston, and L.O. Ramig, "Differential phonatory characteristics of four women with amyotrophic lateral sclerosis," *Journal of Voice*, vol. 8, no. 4, pp. 327–339, 1994.

[20] D. Deliyski, "Acoustic model and evaluation of pathological voice production," in *Third European Conference on Speech Communication and Technology*, 1993.

[21] M. Brückl, "Vocal tremor measurement based on autocorrelation of contours," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[22] M. Brückl, A. Ghio, and F. Viallet, "Measurement of tremor in the voices of speakers with parkinsons disease," *Procedia Computer Science*, vol. 128, pp. 47–54, 2018.

[23] Y. Pantazis, Y. Stylianou, and M. Koutsogiannaki, "A novel method for the extraction of vocal tremor," in *Models and analysis of vocal emissions for biomedical applications: 6th International workshop: December 14-16, 2009, Firenze, Italy*. Firenze University Press, 2009, pp. 1000–1004.

[24] S.O. Sadjadi and J.H.L. Hansen, "Unsupervised speech activity detection using voicing measures and perceptual spectral flux," *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 197–200, 2013.

[25] P. Boersma et al., "Praat, a system for doing phonetics by computer," *Glot international*, vol. 5, 2002.

[26] A. Tsanas, M.A. Little, P.E. McSharry, and L.O. Ramig, "Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average parkinson's disease symptom severity," *Journal of the Royal Society Interface*, vol. 8, no. 59, pp. 842–855, 2011.

[27] N.V. Chawla, K.W. Bowyer, L.O. Hall, and W.P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.