SUBSPACE-BASED FUNDAMENTAL FREQUENCY ESTIMATION

Mads Græsbøll Christensen[†], *Søren Holdt Jensen*[†], *Søren Vang Andersen*[†], *and Andreas Jakobsson*[‡]

[†] Dept. of Communication Technology, Aalborg University, DK-9220 Aalborg, Denmark
Phone: +45 9635 {8672,8654,8627}, email: *{mgc,shj,sva}@kom.auc.dk*
[‡] Dept. of Electrical Engineering, Karlstad University, SE-651 88 Karlstad, Sweden
Phone: +46 (0)54 700 2330, email: *andreas.jakobsson@ieee.org*

ABSTRACT

In this paper, we present a subspace-based fundamental frequency estimator based on an extension of the MUSIC spectral estimator. A noise subspace is obtained from the eigenvalue decomposition of the estimated sample covariance matrix and fundamental frequency candidates are selected as the frequencies where the harmonic signal subspace is closest to being orthogonal to the noise subspace. The performance of the proposed method is evaluated and compared to that of the non-linear least-squares (NLS) estimator and the corresponding Cramér-Rao bound; it is concluded that the proposed method has good statistical performance at a lower computational cost than the statistically efficient NLS estimator.

1. INTRODUCTION

The problem of estimating the fundamental frequency of a periodic signal is a classical problem in signal processing, and throughout the years many different solutions have been suggested to solve it. It is encountered in such applications as, for instance, coding of speech and audio, automatic music transcription and determination of rotating targets in radar. The problem of fundamental frequency estimation can be stated as follows; consider a harmonic signal with the fundamental frequency ω_0 that is corrupted by an additive white complex circularly symmetric Gaussian noise, $w(n)$, i.e.,

$$x(n) = \sum_{l=1}^{L} A_l e^{j(\omega_0 l n + \phi_l)} + w(n), \quad n = 0, \ldots, N-1 \qquad (1)$$

where A_l and ϕ_l are the (real-valued) amplitude and the phase of the l'th harmonic, respectively. The problem considered in this paper amounts to estimating the fundamental frequency ω_0 from a set of N measured samples, $x(n)$. Note that the complex-valued signal model in (1) can also be applied to real-valued signals, when there is little or no spectral contents of interest in the frequencies near 0 and π, by the use of the discrete-time "analytical" signal [1]. The classical fundamental frequency estimators are typically time-domain techniques based on auto-correlation, cross-correlation, the average magnitude difference function (AMDF), or average squared difference function (ASDF). For a historical review of these methods, we refer to [2, 3], and for examples of more recent work we refer to [4, 5, 6]. While subspace techniques, such as the MUltiple SIgnal Classification (MUSIC) algorithm [7], have a rich history in spectral analysis in general, they have only rarely been used in fundamental frequency estimation. In [6, 8], MUSIC is used for finding the individual harmonics independently, and in [9], a noise estimate is obtained from MUSIC and used in a cepstral pitch estimator. In this paper, we propose an extension of the classical MUSIC algorithm by imposing the assumed harmonic structure in (1) on the MUSIC criterion. The paper is organized as follows. In Section 2, the covariance matrix model of the signal model (1) is presented along with some definitions. Then, in Section 3, we present the proposed fundamental frequency estimator termed the harmonically constrained MUSIC estimator. In Section 4, some numerical results are presented and, finally, Section 5 concludes on the work.

This work was conducted within the ARDOR project, EU grant no. IST–2001–34095.

2. COVARIANCE MATRIX MODEL

In this section, we present the covariance matrix model and introduce some useful vector and matrix definitions before we proceed to discuss the proposed extension. By assuming that the phases of the harmonics are independent and uniformly distributed in the interval $[-\pi, \pi]$, the covariance matrix $\mathbf{R} \in \mathbb{C}^{M \times M}$ can be written as [10]

$$\begin{aligned} \mathbf{R} &= \mathrm{E}\left\{\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^H(n)\right\} \\ &= \mathbf{A}(\omega_0)\mathbf{P}\mathbf{A}^H(\omega_0) + \sigma_w^2 \mathbf{I}, \end{aligned} \qquad (2)$$

where $\mathrm{E}\{\cdot\}$ denotes the statistical expectation, $(\cdot)^H$ the conjugate transpose, and $\tilde{\mathbf{x}}(n)$ is a signal vector containing M samples of the observed signal, i.e.,

$$\tilde{\mathbf{x}}(n) = \left[\begin{array}{cccc} x(n) & x(n-1) & \cdots & x(n-M+1) \end{array}\right]^T, \qquad (3)$$

with $(\cdot)^T$ denoting the transpose. Further,

$$\mathbf{P} = \mathrm{diag}\left(\left[\begin{array}{ccc} A_1^2 & \cdots & A_L^2 \end{array}\right]\right) \qquad (4)$$

and the full rank Vandermonde matrix $\mathbf{A}(\omega_0) \in \mathbb{C}^{M \times L}$ is defined as

$$\mathbf{A}(\omega_0) = \left[\begin{array}{ccc} \mathbf{a}(\omega_0) & \cdots & \mathbf{a}(\omega_0 L) \end{array}\right], \qquad (5)$$

where

$$\mathbf{a}(\omega) = \left[\begin{array}{cccc} 1 & e^{-j\omega} & \ldots & e^{-j\omega(M-1)} \end{array}\right]^T. \qquad (6)$$

Also, σ_w^2 denotes the variance of the additive noise, $w(n)$, and $\mathbf{I}$ is the $M \times M$ identity matrix. We note that

$$\mathrm{rank}\left(\mathbf{A}(\omega_0)\mathbf{P}\mathbf{A}^H(\omega_0)\right) = L, \qquad (7)$$

and that the number of harmonics in $\mathbf{A}(\omega_0)$ is bounded by

$$L = \left\lfloor \frac{\omega_{max}}{\omega_0} \right\rfloor, \qquad (8)$$

where ω_{max} may go up to π, although it is typically well below this limit. This is, for example, the case for audio sampled at 44.1 kHz or speech signals sampled at 16 kHz. Here, the constant $M \geq L+1$ is a user parameter that determines the accuracy of the resulting MUSIC frequency estimator, with larger M yielding increasing resolution. Thus, M should be selected as large as possible while still allowing for a reliable estimate of the covariance matrix [10].

3. THE HARMONIC MUSIC ALGORITHM

The MUSIC algorithm [7, 11] (see also [12]) is based on the eigenvalue decomposition (EVD) of the covariance matrix $\mathbf{R}$, exploiting the structure in (2). Let

$$\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^H \tag{9}$$

where $\mathbf{U}$ is formed from the M orthonormal eigenvectors of $\mathbf{R}$, i.e.,

$$\mathbf{U} = [\ \mathbf{u}_1 \quad \cdots \quad \mathbf{u}_M\ ], \tag{10}$$

and Λ is a diagonal matrix with the eigenvalues, λ_k, on the diagonal. The following decomposition requires *a priori* knowledge of the number of harmonic components, L. Here, we will instead determine L using (8), and as a result this L will be frequency dependent; hereafter, we will use the notation $L(\omega_0)$ to stress this dependence. Now, let $\mathbf{G}(\omega_0)$ be formed from the $M-L(\omega_0)$ eigenvectors corresponding to the $M-L(\omega_0)$ least significant eigenvalues ($\mathbf{G}(\omega_0)$ is a function of ω_0 through $L(\omega_0)$), i.e.,

$$\mathbf{G}(\omega_0) = \begin{bmatrix} \mathbf{u}_{M-L(\omega_0)+1} & \cdots & \mathbf{u}_M \end{bmatrix}. \tag{11}$$

Then, it can be shown that the noise subspace spanned by $\mathbf{G}(\omega_0)$ will be orthogonal to the Vandermonde matrix $\mathbf{A}(\omega_0)$ spanned by the L harmonic sinusoids [10], i.e.,

$$\mathbf{A}^H(\omega_0)\mathbf{G}(\omega_0) = \mathbf{0}. \tag{12}$$

We stress that where $\mathbf{A}$ is a function of the set of frequencies $\{\omega_l\}_{l=1}^{L}$ in MUSIC, it is here only a function of the fundamental frequency ω_0 as the frequencies of the harmonics are given by $\omega_l = \omega_0 l$. As $\mathbf{R}$ is typically unknown, one needs to form an estimate of it; here, we estimate the sample covariance matrix as

$$\widehat{\mathbf{R}} = \frac{1}{N}\sum_{n=M}^{N}\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^H(n). \tag{13}$$

and note that the orthogonality in (12) will only hold approximately for the eigenvectors found from this matrix. Exploiting the harmonic structure in (1), the estimated fundamental frequency can be found as

$$\arg\min_{\omega_0} \|\mathbf{A}^H(\omega_0)\mathbf{G}(\omega_0)\|_F, \tag{14}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. By the Cauchy-Schwarz inequality, we have that

$$\|\mathbf{A}^H(\omega_0)\mathbf{G}(\omega_0)\|_F \leq \|\mathbf{A}^H(\omega_0)\|_F\|\mathbf{G}(\omega_0)\|_F. \tag{15}$$

As the $M-L(\omega_0)$ columns of $\mathbf{G}(\omega_0)$ are orthonormal, and all the $L(\omega_0)$ columns of $\mathbf{A}(\omega_0)$ have norm $\sqrt{M}$, we get

$$\|\mathbf{A}^H(\omega_0)\mathbf{G}(\omega_0)\|_F \leq \sqrt{L(\omega_0)M}\sqrt{M-L(\omega_0)} \tag{16}$$

and thus

$$\frac{\|\mathbf{A}^H(\omega_0)\mathbf{G}(\omega_0)\|_F}{\sqrt{L(\omega_0)M(M-L(\omega_0))}} \leq 1. \tag{17}$$

We now define the harmonic pseudo-spectrum as

$$P(\omega_0) = \frac{L(\omega_0)M(M-L(\omega_0))}{\|\mathbf{A}^H(\omega_0)\mathbf{G}(\omega_0)\|_F^2}, \tag{18}$$

and find the estimated fundamental frequency as

$$\hat{\omega}_0 = \arg\max_{\omega_0\in\Omega_0} P(\omega_0). \tag{19}$$

Thus, the fundamental frequency candidates can be found from (18) by sweeping ω_0 over a finite set of frequencies $\mathbf{\Omega}_0$ and then project the harmonic subspace onto the noise subspace. In the rest of this paper, we refer to this estimator as the *harmonically constrained* MUSIC or HMUSIC in short. The algorithm can be summarized as the following steps:

1. Estimate $\widehat{\mathbf{R}}$ using (13).
2. Perform an EVD of $\widehat{\mathbf{R}}$.
3. For each $\omega_0 \in \Omega_0$,
   (a) Determine $L(\omega_0)$ from (8).
   (b) Construct $\mathbf{A}(\omega_0)$ using (5), and $\mathbf{G}(\omega_0)$ using (11).
   (c) Compute $P(\omega_0)$ using (18).
4. Find fundamental frequency candidates as the maxima of $P(\omega_0)$.

We note that it is possible to use a noise subspace with a fixed dimension by estimating an upper bound on L. For example, in speech the fundamental frequency is typically limited to the range 60 Hz - 400 Hz, which would result in an upper bound of the dimension of the signal subspace

$$L = \left\lfloor \frac{\omega_{max} f_s}{2\pi 60} \right\rfloor, \tag{20}$$

with f_s being the sampling frequency. In our experience, the peaks of the harmonic pseudo-spectra computed using a fixed L are often more distinct and thus appear less noisy compared to the variable dimension approach, but the latter seems to give a better response at low frequencies.

A classical problem in fundamental frequency estimation is erroneous estimates at k or $1/k$ times the true fundamental frequency for $k = 2, 3, \ldots$, commonly referred to as doublings and halvings. These problems also exist in HMUSIC. Especially, doublings of the fundamental occur, because $\mathbf{A}(k\omega_0)$, for $k = 2, 3, \ldots$, is spanned by the column space of $\mathbf{A}(\omega_0)$ and these columns are thus also orthogonal to the noise subspace when this is kept fixed. This is less of a problem for variable dimension noise subspace. Halvings of the fundamental frequency also occur, but these are generally much weaker than the doublings as only a subset of the harmonics will be orthogonal to the noise subspace. In order to build a practical fundamental frequency estimator from HMUSIC, we need to limit the search space $\mathbf{\Omega}_0$ of ω_0 to some interval that does not include doublings and halvings. This can, for example, be achieved by pitch tracking, some coarse initial estimate, or by some post-processing of the harmonic pseudo-spectrum. In this paper, we defer from any further discussion of this and instead concentrate on the statistical performance of the estimator.

4. EXPERIMENTAL RESULTS

4.1 Reference Methods

For reference, we use a non-linear least-squares (NLS) fundamental frequency estimator similar to that of [6]. This is a particularly simple version of the NLS frequency estimator (see, e.g., [10]) because of the harmonic relation between the sinusoidal components. As is well known, the NLS frequency estimator is statistically efficient under white noise conditions; furthermore, it can be shown that the NLS estimator is asymptotically efficient also for the coloured noise case [13]. We note that the NLS fundamental frequency estimator can be stated as the minimizer of the squared error between the signal and the harmonic sinusoidal model, and be found by sweeping over a finite set Ω_0 of frequencies. Here, we use the same grid as in HMUSIC. We refer to this method as harmonically constrained NLS (HNLS). While the HMUSIC gives strong false peaks at integer multiples of the fundamental frequency, the HNLS estimator is very prone to halvings ($1/k$ with $k = 2,3,\ldots$) because a fundamental frequency of $0.5\omega_0$ will capture more signal energy than the true fundamental, especially under noisy conditions. Thus, like the HMUSIC, we need to limit the search range Ω_0 in order to get the correct result. For each grid point in Ω_0, HNLS is computationally more complex than HMUSIC as HNLS involves a matrix inversion and matrix products whereas HMUSIC involves only a matrix product for each frequency point. There is, however, some additional computational overhead associated with HMUSIC as it requires the calculation of the sample covariance matrix and an EVD. As the resolution of the grid increases, the relative influence of this overhead decreases. As an additional reference, we also use spectral MUSIC [7, 11] on the same grid as HMUSIC and HNLS to locate the frequency of the first harmonic. This method does not take the harmonic structure of the spectrum into account.

4.2 Speech Signal

In this section, we show harmonic pseudo-spectra of a speech signal (female speaker, sampled at 8 kHz) and illustrate the difference between using a fixed dimensional noise subspace and a variable dimensional one. In Figure 1(b), the harmonic pseudo-spectrum of the segment of voiced speech in Figure 1(a) is depicted. This pseudo-spectrum has been calculated using a fixed noise subspace in the sweep over ω_0. It can be seen that the fundamental frequency stands out very clearly at approximately 159 Hz and that the double is very noticeably present at 318 Hz. As a comparison, the harmonic pseudo-spectrum with a variable dimension noise subspace is shown in Figure 1(c), clearly illustrating the reduced risk for a pitch-doubling. It can also be seen that the peaks of Figure 1(b) are more distinct compared to the noise floor than those of 1(c).

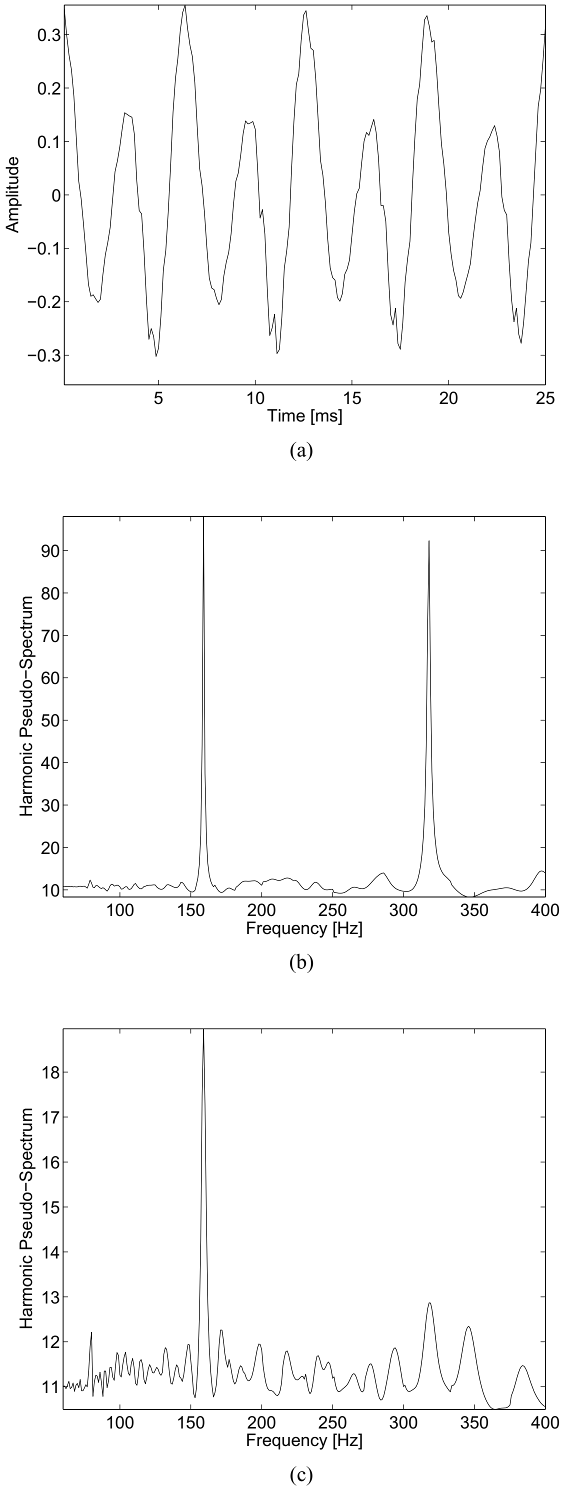

Figure 1: (a) Voiced speech segment, (b) harmonic pseudo-spectrum of speech segment using a fixed noise subspace, and (c) using a variable dimension subspace.

4.3 Synthetic Signals

To investigate the statistical efficiency of HMUSIC, we perform an evaluation of the fundamental frequency estimator on a synthetic signal using a technique similar to those of [6, 14]. As a comparison, we also show the asymptotic Cramér-Rao bound (CRB) as derived in [14]. First, we investigate the effects of varying *SNR* for a fixed segment length. The *SNR* is defined as $SNR = 10\log_{10}(\sigma_s^2/\sigma_w^2)$, with σ_s^2 being the variance of the sinusoidal part of (1) and σ_w^2 being the variance of the noise. In Figure 2, the standard devision of MUSIC, HNLS, HMUSIC and the CRB are shown as a function of the *SNR* for a segment length of 256 samples. These were found by 200 Monte Carlo simulations, where in

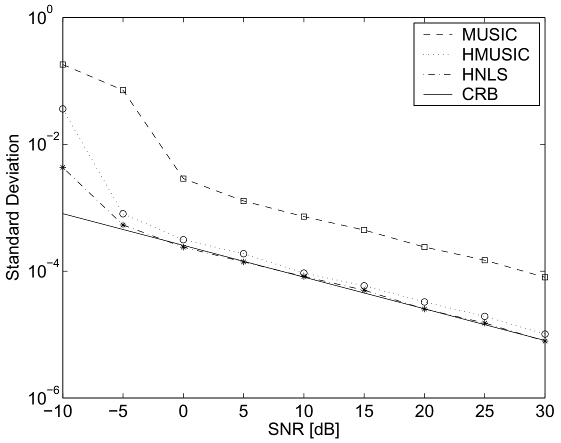

Figure 2: Standard deviation of the estimates $\hat{\omega}_0$ and the CRB for varying *SNR* for $N = 256$.

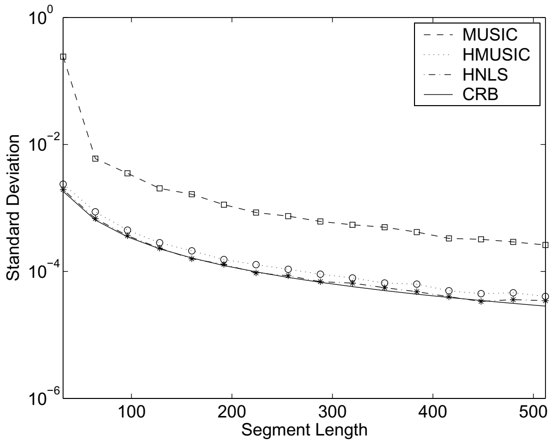

Figure 3: Standard deviation of the estimates $\hat{\omega}_0$ and the CRB for varying segment lengths N with $SNR = 10$dB.

each run the additive noise sequence and the phases of the harmonics have been randomized. A fundamental frequency of $\omega_0 = 2\pi 0.08$ (corresponding to 640 Hz at 8 kHz sampling frequency) was used in all simulations. Five real harmonics ($L = 10$) were used, all with an amplitude of 1, and the step-size of the grid searches of MUSIC, HNLS and HMUSIC was set to 0.01 Hz. Further, Ω_0 was constrained to be in the vicinity of ω_0 by $\pm 10\%$, and M was set to 128.

The effects of varying segment lengths, N, for a fixed SNR have also been investigated. The results are shown in Figure 3 for an SNR of 10 dB. Here a stepsize of 0.1 Hz was used and the dimensions of the sample covariance matrix was set to $M = \lfloor N/2 \rfloor$. Note that the HMUSIC and MUSIC algorithms are sensitive to the choice of M relative to N. From these figures, it can be seen that HMUSIC has very good statistical performance approaching the Cramér-Rao bound. From observing the performance of HMUSIC compared to MUSIC, it can also be seen that there is a big gain in taking the harmonic structure into account in the estimation.

5. CONCLUSION

In this paper, a subspace-based fundamental frequency estimator has been proposed. This estimator is based on a harmonic extension of the classical MUSIC estimator, letting the dimensionality of the noise signal subspace depend on the underlying fundamental frequency. The resulting estimator is obtained by sweeping over a set of frequencies. The performance of the estimator has been evaluated and compared to both the non-linear least-squares estimator, the classical MUSIC algorithm, and the Cramér-Rao bound. From the simulations, we conclude that the estimator has good statistical performance at a computational complexity, which is lower than the nonlinear least-squares for high resolutions.

REFERENCES

[1] S. L. Marple, “Computing the discrete-time "analytic" signal via FFT,” in *IEEE Trans. Signal Processing*, Sept. 1999, vol. 47.

[2] W. Hess, *Pitch Determination of Speech Signals*, Springer-Verlag, Berlin, 1983.

[3] W. Hess, “Pitch and voicing determination,” in *Advances in Speech Signal Processing*, S. Furui and M. M. Sohndi, Eds., pp. 3–48. Marcel Dekker, New York, 1992.

[4] A. de Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” in *Journal of ASA*, Apr. 2002, vol. 111(4).

[5] D. E. Terez, “Robust pitch determination using nonlinear state-space embedding,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 2002.

[6] H. Li, P. Stoica, and J. Li, “Computationally efficient parameter estimation for harmonic sinusoidal signals,” in *Signal Processing*, 2000, vol. 80, pp. 1937–1944.

[7] R. O. Schmidt, “Multiple emitter location and signal parameter estimation,” in *IEEE Trans. Antennas and Propagation*, Mar. 1986, vol. 34(3).

[8] N. Malik and W. H. Holmes, “Pitch estimation and a measure of voicing from pseudo-spectra,” in *Fifth International Symposium on Signal Processing and its Applications*, 1999.

[9] M. S. Andrews, J. Picone, and R. D. Degroat, “Robust pitch determination via SVD based cepstral methods,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1990.

[10] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice Hall, 1997.

[11] G. Bienvenu, “Influence of the spatial coherence of the background noise on high resolution passive methods,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, 1979.

[12] G. H. Händel, “On the History of MUSIC,” in *IEEE Signal Processing Magazine*, March 1999, p. 13.

[13] P. Stoica, A. Jakobsson, and J. Li, “Cisiod parameter estimation in the coloured noise case: Asymptotic Cramér-Rao bound, maximum likelihood, and nonlinear least-squares,” in *IEEE Trans. Signal Processing*, Aug. 1996, vol. 45(8).

[14] A. Nehorai and B. Porat, “Adaptive comb filtering for harmonic signal enhancement,” in *IEEE Trans. Antennas and Propagation*, Oct. 1986, vol. 34(5).