Rate Allocation for Non-Collaborative Multi-User Speech Communication Systems Based On Bargaining Theory

Bengt J. Borgström, *Student Member, IEEE,* Mihaela van der Schaar, *Senior Member, IEEE,*
and Abeer Alwan, *Senior Member, IEEE*

Abstract— We propose a novel rate allocation algorithm for multi-user speech communication systems based on bargaining theory. Specifically, we apply the generalized Kalai-Smorodinsky bargaining solution since it allows varying bargaining powers to match the dynamic nature of speech signals. We propose a novel method to derive bargaining powers based on the short-time energy of the input speech signals, and subsequently allocate rates accordingly to the users. An important merit of the proposed framework is that it is general and can be applicable for resource allocation across a variety of multi-rate speech coders and it is robust to a variety of speech quality metrics. The proposed system is also shown to involve a quick and low complexity training process. We generalize the algorithm to scenarios in which users have unequally weighted priorities. These scenarios might arise in emergency situations, in which certain users are more important than others. The proposed rate allocation system is shown to increase the utility measures for both the Itakura and segmental SNR functions relative to the baseline system that performs uniform rate allocation. Additionally, although the instantaneous bitrate resolution of the speech encoder is not changed, the proposed system is shown to increase the short-time average bitrate resolution, and therefore provides a greater number of operational rate modes for the network.

I. INTRODUCTION

Resource allocation techniques in multi-user communication systems have been a major topic of research for many decades. Specifically, thorough attention has been paid to dynamic rate allocation within cellular telephone networks due to the inherent limits on resources in such networks. However, with the emergence of new modes of communication, such as Internet phone services, rate allocation algorithms for speech transmission systems have again become a crucial topic of research.

This paper addresses the problem of dynamic rate allocation across multiple non-collaborative speech communication systems. Dynamic resource allocation has been studied as an efficient alternative to fixed resource allocation in general wireless data networks [1] [2]. However, these techniques consider only spectral allocation, and determine solutions based on external factors such as time-varying loads and spatial channel differences. In [3], the authors propose a rate control approach for generalized processor sharing. The method parameterizes a source model and allocates rate accordingly, but it does not consider resulting utilities for multimedia applications.

Attempts have been made to solve resource allocation problems in the utility domain by considering the quality of service (QoS) to users. In [4], dynamic spectral allocation is based on QoS requirements in terms of bit error rate (BER), but the proposed method does not consider the results on the decoded data. Other existing resource allocation schemes based on utility measures often require a relatively simple utility function. In [5], the authors propose a rate allocation scheme for communication networks based on utility results, but the scheme requires the utility function to be solvable with Lagrangian optimization techniques.

Recent research has applied game theory and bargaining theory to solve resource allocation problems in the utility domain, and these techniques have been shown to provide improved performance. In [6], game theory is applied to power control in cellular systems. In [7], the Nash bargaining solution is used to develop an auctioning algorithm for channel allocation in wireless networks, and in [8], bargaining theory is applied to rate allocation in multi-user video transmission systems. Also, note that none of the previously mentioned studies develop resource allocation algorithms based on speech characteristics, and instead base rate allocation on channel and/or network conditions.

In this paper, we propose a non-collaborative rate allocation algorithm based on axiomatic bargaining theory for transmission of speech signals. In order to apply axiomatic bargaining theory to the rate allocation problem, this paper defines certain aspects of the general normalized bargaining problem in terms of speech processing and communication theory. Specifically, this paper focuses on a multi-user system with a central spectral moderator (CSM), which is responsible for dynamically and fairly allocating rate to users in the network.

The novelty of this paper lies in the fact that we apply the Kalai-Smorodinsky bargaining solution to speech communication networks. Various resource allocation solutions based on bargaining theory can be adopted to enable a fair division of resources among users, such as the Nash bargaining solution [9] and the Raiffa bargaining solution [10]. However, the Kalai-Smorodinsky bargaining solution is especially useful for multi-user speech communication as it allocates the resources in such a way that the achieved utility of any participating user results in the same quality penalty, i.e. the same decrease in speech quality, of the other users relative to their maximum achievable qualities.

Additionally, the Kalai-Smorodinsky bargaining solution allows the use of bargaining powers to weight users differently in time. We introduce a novel method for deriving bargaining

powers based on the short-time energy of the input speech signals to match the dynamic nature of speech.

Unlike conventional solutions like Lagrangian optimization, the Kalai-Smorodinsky bargaining solution does not require the utility-resource function to fulfill specific properties such as convexity. Importantly, this solution does not even require to have an analytical expression for the utility-resource tradeoffs, which is essential for speech coders.

Another important feature of the proposed system is the robustness of the algorithm to different speech coders and different speech quality metrics. The rate allocation algorithm derived in this paper is compatible with any speech encoder, as long as it can produce bitstreams at multiple quality levels. The proposed system is also compatible with any numeric speech quality measure, which is important since there is no generally favored numerical quality of speech metric.

In Section II, we review several bargaining theory fundamentals. Section III describes the application of bargaining theory to the problem of rate allocation, and develops the proposed rate allocation system. Section IV analyzes the complexity of the algorithm, and describes the learning process of the system. Section V shows the performance and results of the overall system. Finally, conclusions and discussion are provided in Section VI.

II. REVIEW OF BARGAINING THEORY FUNDAMENTALS

A. Axiomatic Bargaining Theory

A bargaining problem involves two or more users who can collaborate for their mutual benefit in multiple ways [9]. A bargaining solution is defined as an optimal distribution of resources among the users involved. A bargaining solution must lie on the Pareto surface, which is defined as the collection of points in the utility space relative to which no other solutions are superior in all objectives [11].

The development of Axiomatic Theory of Bargaining in [9] introduced a more mathematical approach to the solution. In axiomatic bargaining theory, a solution is selected that satisfies a set of rational and desirable axioms, and these axioms guarantee fairness among parties. Specifically, this theory presented the *normalized* bargaining problem, in which the problem is represented by the pair $(S, \mathbf{d})$ in utility space. In this notation, S is the subset of the utility space which includes all feasible utility points, and $\mathbf{d}$ is the *disagreement point* defined as:

$$\mathbf{d} = \left[u_1^{min}, u_2^{min}, \ldots, u_M^{min} \right]^T, \qquad (1)$$

where u_i^{min} represents the minimum agreeable utility for user i.

B. The Kalai-Smorodinsky Bargaining Solution

The Kalai-Smorodinsky Bargaining Solution (KSBS) [12] is confined to functions $f : R^M \to R^M$ such that $f(S, \mathbf{d}) \in S$ that satisfy the following three axioms which are introduced in [9]. Note that we define the vector operator $>$ such that for $\mathbf{x}, \mathbf{y} \in R^M$, $\mathbf{x} > \mathbf{y}$ iff $x_i > y_i$ for $i = 1, 2, \ldots, M$.

Axiom 1- *Pareto Optimality*: For every $(S, \mathbf{d})$ there is no $\mathbf{y} \in S$ such that $y > f(S, \mathbf{d})$ and $y \neq f(S, \mathbf{d})$.

Axiom 2- *Symmetry*: We let $T : R^M \to R^M$ be defined by $T((x_1, x_2, \ldots, x_M)) = (x_{p(1)}, x_{p(2)}, \ldots, x_{p(M)})$, where $\mathbf{p}$ is any permutation of the numbers $(1, \ldots, M)$, and we require that for every $(S, \mathbf{d}) \in R^M$, $f(T(S), T(\mathbf{d})) = T(f(S, \mathbf{d}))$.

Axiom 3- *Invariance with Respect to Affine Transformation of Utility*: A is an affine transformation of utility if $A = (A_1, A_2, \ldots, A_M) : R^M \to R^M$, $A(x_1, x_2, \ldots, x_M) = (A_1(x_1), A_2(x_2), \ldots, A_M(x_M))$, and the maps $A_i(x)$ are of the form $c_i x + d_i$ for some positive constant c_i and some constant d_i. We require that for such a transformation A, $f(A(S), A(\mathbf{d})) = A(f(S, \mathbf{d}))$.

These axioms guarantee fairness and efficiency of the bargaining solution. Specifically, *Axiom 1* ensures that the bargaining solution is efficient and that no other solution can be found that can assign a greater amount of utility to all users. *Axiom 2* guarantees that users are given equal treatment with respect to assignment of utility. Finally, *Axiom 3* guarantees that if users are bargaining for unlike utilities, these utilities will be normalized before the bargaining solution is determined.

In [12], a fourth axiom is introduced, namely the *Axiom of Monotonicity*. This axiom states in the 2-user case that if user 1 demands a certain utility level, and the utility level of user 2 can simultaneously be increased, then the utility level of user 2 assigned by the solution should be increased.

Let us define the function $g_S(x_i)$ as the maximum utility levels other users can be assigned if user i is assigned utility level x_i. Then, the *Axiom of Monotonicity* is defined as:

Axiom of Monotonicity- If $(S_2, \mathbf{d})$ and $(S_1, \mathbf{d})$ are bargaining pairs such that $g_{S_1} < g_{S_2}$, then $f_i(S_1, \mathbf{d}) < f_i(S_2, \mathbf{d})$ (where $f(S, \mathbf{d}) = (f_1(S, \mathbf{d}), \ldots, f_M(S, \mathbf{d}))$).

It is proven in [12] that there is one and only one solution within the set of functions defined by *Axioms 1,2,* and *3*, that also satisfies the *Axiom of Monotonicity*. This unique KSBS introduced the concept of a *utopia point*, $\mathbf{a}$, which is a point in the utility space defined as:

$$\mathbf{a} = [u_1^{max}, u_2^{max}, \ldots, u_M^{max}]^T, \qquad (2)$$

where u_i^{max} represents the maximum achievable utility for user i and is only achievable if all the rate is allocated to that user.

The KSBS defines the unique solution to a normalized bargaining problem as the intersection of the Pareto Surface and the line connecting $\mathbf{d}$ and $\mathbf{a}$. The KSBS also offers the ability to apply the concept of *bargaining powers*, which represents the weights of users' demands in bargaining.

The unique KSBS, denoted as $\widehat{\mathbf{u}}$, to the bargaining problem represented by $(S, \mathbf{d})$ [12], is defined as:

$$\widehat{\mathbf{u}}_k = \max_{\mathbf{u}_k \in S, \ \alpha \in R} \mathbf{u}_k, \qquad (3)$$
$$\text{where } \mathbf{u}_k = \mathbf{d} + \alpha C \cdot (\mathbf{a} - \mathbf{d})$$

where C is a diagonal matrix whose c_{ii} element is the bargaining power of the ith user, and α is a scalar. Figure
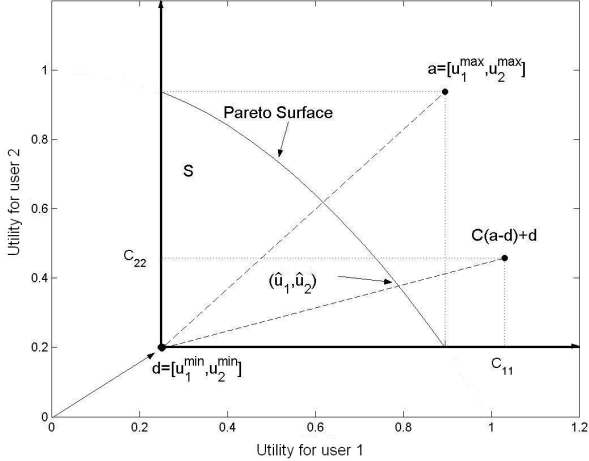
Fig. 1. *The Kalai-Smorodinsky Bargaining Solution (KSBS) for the 2-user case*

TABLE I
Operating Modes of the GSM AMR-NB Speech Encoder, where u_{SNR} and u_{Itak} Represent the Segmental SNR and Inverse Itakura Utility Functions Respectively

Mode	Bitrate (in kbps)	u_{SNR}	u_{Itak}
1	4.75	1.541	2.912
2	5.15	1.385	3.020
3	5.90	1.610	3.557
4	6.70	1.661	3.726
5	7.40	2.244	3.916
6	7.95	1.881	3.864
7	10.2	2.783	5.278
8	12.2	2.957	5.173

1 shows an example of the KSBS, labeled as $(\widehat{u}_1, \widehat{u}_2)$, in the simple 2-user case.

III. PROPOSED RATE ALLOCATION SYSTEM

A. System Overview

Application of bargaining theory to the problem of rate allocation necessitates certain definitions. Firstly, the concept of resources can clearly be defined as allocated rate to each user in the system. Additionally, the concept of utility can be defined as the quality of speech transmitted by each user. Finally, the concept of bargaining power can be interpreted as the relative benefit of additional rate for the speech quality of user's transmitted speech.

As mentioned in the introduction, this paper will focus on a multi-user speech communication system with a CSM for the M-user case. The CSM is responsible for dynamically allocating a constant total rate, R, to the users. That is, for every block of coded speech, the CSM will determine the rate allocation vector:

$$\mathbf{q}_k = [r_{1,k}, r_{2,k}, \ldots, r_{M,k}]^T, \quad (4)$$

where $r_{i,k}$ represents the rate allocated to the kth block from user i, for $1 \leq i \leq M$ and $k \geq 1$. Note the following constraints on $\mathbf{q}_k$:

- $0 \leq r_i \leq R$, for $1 \leq i \leq M$ and $k \geq 1$
- $\sum_{i=1}^{M} r_{i,k} \leq R$, for $k \geq 1$
- The value $\Delta R = R - \sum_{i=1}^{M} r_{i,k}$ is less than the amount of rate necessary for any user to operate at a higher bitrate mode of the encoder. That is, the available rate R has been maximally allocated.

Figure 2 shows the system overview for the 2-user case. In this figure, $s_i(n)$ represents the input speech signal of user i and $BP_{i,k}$ represents the bargaining power of user i for block k. Also, $B_{i,k}$ represents the kth block of coded speech for user i.

In our proposed system, the CSM determines the solution rate allocation vector $\widehat{\mathbf{q}}_k$ through the following steps:

1) Define the Pareto Surface (see Section III-C.1).
2) Position the disagreement point (see Section III-C.2).
3) Determine the utopia point (see Section III-C.3).
4) Determine the matrix C by normalizing the bargaining powers of the users (see Section III-C.4).
5) Determine $\widehat{\mathbf{u}}$, the KSBS in the utility space (see Section III-C.6).
6) Determine the point in the resource domain, $\widehat{\mathbf{q}}_k$, corresponding to $\widehat{\mathbf{u}}_k$ (see Section III-C.7). Note that $\widehat{\mathbf{q}}_k$ is an approximation of $\mathbf{q}_k$, since the Pareto Surface is discrete, and the exact operating point $\mathbf{q}_k$ can not generally be achieved.

As illustrated in Figure 2, the computational steps of the proposed rate allocation system occur at two different locations within the system. The calculations of the unnormalized bargaining powers are carried out within individual speech encoder blocks, and therefore these calculations can be done in parallel. The unnormalized bargaining powers are then received by the CSM, and the CSM determines the KSBS. The components of the solution rate vector are then transmitted to the corresponding speech encoder blocks.

The rate allocation system was designed to be updated for every block of coded speech. Thus, in the case of most modern speech coders, a new rate allocation solution is determined approximately every 20 ms. The proposed algorithm is therefore able to adapt to transitions in the speech signals as quickly as the coder can adapt.

It is important to note that the rate allocation algorithm developed in this paper can be applied to many different speech encoders. As can be seen in Figure 2, the only requirement for the coder used is that it can operate at multiple levels of quality. Furthermore, better performance of our rate allocation scheme can be expected with an increased number of quality levels. This is due to the fact that an increased number of operating modes of the speech encoder will provide increased resolution on the Pareto Surface. To illustrate the performance of our system, we use the GSM Adaptive Multi-Rate (AMR) Narrow-Band (NB) speech coder specified in [13]. This speech coder can operate in 8 different rate modes between 4.75 kbps and 12.2 kbps.
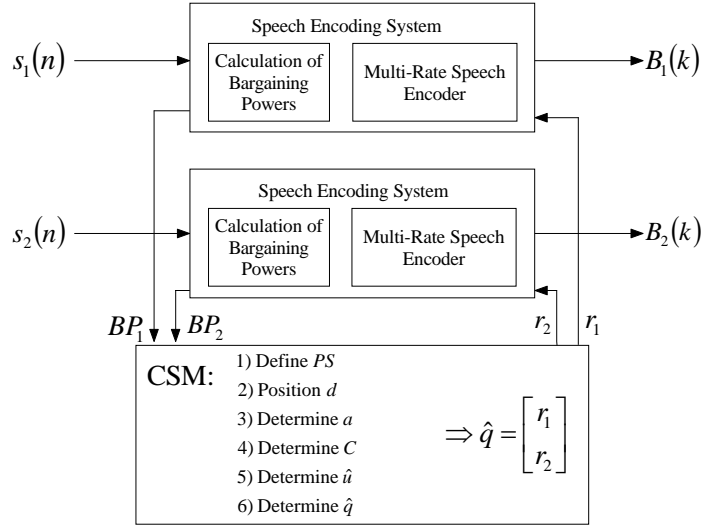
Fig. 2. *System Overview for 2-User Case where $s_i(n)$ Represents the Input Speech Signal of User i, BP_i Represents the Unnormalized Bargaining Power of User i, r_i Represents the Rate Allocated to User i, and $B_i(k)$ Represents the kth Block of Coded Speech From User i*

B. Definitions of the Utility Functions and Feasibility Set

In general bargaining theory, the utility function is a function from the resource domain to the utility domain. In the specific case of our system, the utility function expresses the quality of synthesized speech encoded at a certain bit rate.

To illustrate the robustness of our system to a variety of speech quality metrics, we define two distinct utility functions. These utility functions are based on common speech distortion measures.

A widely used objective speech quality metric for synthesized speech is the segmental SNR, d_{SNR} [16]. If the original speech signal, $s(n)$, is encoded at a bit rate r, the segmental SNR of the synthesized speech signal, $\widehat{s}_r(n)$, is defined as:

$$d_{SNR}(r) = \frac{10}{K}\left[\sum_{k=0}^{K-1}\log_{10}\left\{\sum_{n=kN}^{(k+1)N-1}\frac{s^2(n)}{(s(n)-\widehat{s}_r(n))^2}\right\}\right],$$
$$(5)$$

where N is the frame size and K is the total number of frames. The segmental SNR utility function is therefore given by:

$$U_{SNR}(\mathbf{q}_k) = [d_{SNR}(r_{1,k}), d_{SNR}(r_{2,k}), \ldots, d_{SNR}(r_{M,k})].$$
$$(6)$$

We also consider another distortion metric which is often used to compare Linear Prediction Coding (LPC) based coefficients in speech coding and recognition applications. [15]. The Itakura distortion, $I_{Itak}(r)$, between the original speech signal, $s(n)$, and synthesized speech encoded at a bit rate r, $\widehat{s}_r(n)$, is defined as:

$$I_{Itak}(r) = \log\left(\frac{\widehat{\mathbf{a}}^T R_p \widehat{\mathbf{a}}}{\mathbf{a}^T R_p \mathbf{a}}\right) \qquad (7)$$

where R_p is the pth-order autocorrelation matrix of $s(n)$, and $\mathbf{a}$ and $\widehat{\mathbf{a}}$ are defined as:

$$\mathbf{a} = [1, -a_1, -a_2, \ldots, -a_p]^T, \qquad (8)$$

and

$$\widehat{\mathbf{a}} = [1, -\widehat{a}_1, -\widehat{a}_2, \ldots, -\widehat{a}_p]^T \qquad (9)$$

where a_k and $\widehat{a}_k$ are the kth predictor coefficients for p-order Linear Predictive Coding (LPC) [17] analysis of $s(n)$ and $\widehat{s}(n)$, respectively.

In order to obtain a utility function that is directly related to speech quality, we define our proposed utility function as the inverse of the Itakura distortion. Thus, our utility function is given by:

$$U_{Itak}(\mathbf{q}_k) = \left[\frac{1}{I_{Itak}(r_{1,k})}, \frac{1}{I_{Itak}(r_{2,k})}, \ldots, \frac{1}{I_{Itak}(r_{M,k})}\right].$$
$$(10)$$

Table I shows the operating modes and resulting utility measures of the GSM AMR-NB speech encoder. The utility functions used are the Segmental SNR distortion [16] and the inverse of the Itakura distortion [15] previously described. The utility points are obtained using a training set from the TIMIT Acoustic-Phonetic Continuous Speech Corpus [19]. The training set included 275 sentences, 169 of which were spoken by various male speakers, and 106 of which were spoken by various female speakers.

Note that there are a couple of instances where the utility points decrease with an increase in bitrate. These instances are due to the varying bit allocation in the speech coding specifications of the GSM AMR-NB encoder. For example, the drop in the u_{SNR} utility from mode 1 to mode 2 is most likely due to the fact that the mode 1 coded speech block contains 2 sets of gain values per 20 ms block, and designates 8 bits to each. The mode 2 block contains 4 sets of gain values, but only designates 6 bits to each. Also, the drop in performance for both utility functions in mode 6 is most likely due to the

difference in adaptive codebook construction. Finally, the drop in the u_{Itak} utility from mode 7 to mode 8 is most likely due to the fact that the mode 8 block uses no look-ahead for calculation of the the linear prediction coefficients.

The quality feasibility set is defined as the set of all points in the utility space which correspond to feasible resource allocation vectors. Specifically, the quality feasibility set is defined as:

$$S = \left\{ \mathbf{q} \mid \exists \mathbf{r} = [r_1, r_2, \ldots, r_M]^T, \quad \mathbf{r} \text{ is feasible,} \right.$$
$$\left. \text{and } u_i^{min} \leq u_i(r_i) = q_i, \ i = 1, 2, \ldots, M \right\}. \quad (11)$$

An important property required to be able to apply the KSBS to our problem of rate allocation is the fact that the quality feasibility set is **d**-*comprehensive*. The definition of comprehensiveness of a set states that the set $S \subset R^M$ is **d**-*comprehensive* if $y \in S$ and $d \leq x \leq y$ imply $x \in S$.

Lemma 1: The quality feasibility set S is **d**-comprehensive.
Proof: Let $\mathbf{x}, \mathbf{y} \in R^{M \times 1}$ be vectors in the feasibility set such that $\mathbf{d} \leq \mathbf{x} \leq \mathbf{y}$. Now define $\mathbf{r}_y \in R^{M \times 1}$ such that $\mathbf{y} = U(\mathbf{r}_y)$. Since our utility function is monotonically increasing, there must exist a vector $\mathbf{r}_x \in R^{M \times 1}$ such that $\mathbf{x} = U(\mathbf{r}_x)$.

C. Definitions of Bargaining Theory Parameters for Proposed System

1) Determining the Pareto Surface: To ensure that a bargaining solution is not wasteful, a fundamental property of a bargaining solution is Pareto optimality. A solution $\mathbf{q}$ is Pareto optimal if $\mathbf{q} \in PS$, where PS is the Pareto Surface of S and is defined as:

$$PS = \{\mathbf{x} \in S \mid \mathbf{y} > \mathbf{x} \text{ implies } \mathbf{y} \notin S\}. \quad (12)$$

Figure 3 illustrates an example of the Pareto Surface using U_{SNR} for a total rate constraint of 18.0 kbps in the 2-user case, and the corresponding user rates are shown. Note that the Pareto Surface in Figure 3 includes few points. This is due to the fact that the GSM AMR-NB encoder can operate in a small number of quality modes across a large range of bitrates. A speech encoder with a larger number of operating modes with finer bitrate resolution would result in a more populated Pareto Surface.

In the proposed system, the Pareto Surface is obtained by first compiling a list of all possible operating points for the M-user case. Note that these operating points, given as $(r_1, r_2, \ldots, r_M)$, are comprised of user rates that lie between the rates corresponding to the utility values u_i^{min} and u_i^{max}. Thus, the system contains a set of predetermined lists for different user numbers. The Pareto Surface for a given rate constraint R, and a given number of users M, is then obtained through the following steps:

- The list corresponding to the M-user case is traversed, and the total rate is determined for each operating point: $R_T = \sum_{m=1}^{M} r_m$. If $R_T > R$, the point is excluded from the subset of possible Pareto Surface points.
- The remaining list of possible operating points is traversed, and each point is checked for Pareto Optimality.
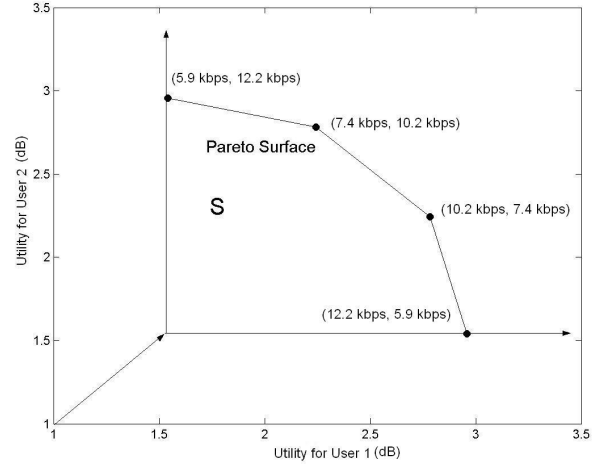

Fig. 3. *Example of the Pareto Surface for the 2-User Case Using U_{SNR} With Total Rate Constraint of 18.0 kbps*

That is, for the current point, $\mathbf{u}_{curr}$, and for any other point $\mathbf{x}$, if $\mathbf{u}_{curr} < \mathbf{x}$, the current point is excluded.

The computational load of obtaining the Pareto Surface for a given R and M may become large as M grows large. However, these algorithms are carried out offline, and the Pareto Surface is obtained for the particular rate constraint and user number before the system is used.

2) Determining the Disagreement Point: As shown in Equation 1, the disagreement point, $\mathbf{d}_k$, is the point in utility space composed of the minimum acceptable utility levels for each user. Thus, $\mathbf{d}_k$ represents the point in utility space below which solutions are deemed unacceptable by one or more users. Only utility points greater than $\mathbf{d}_k$ are therefore considered during the bargaining process.

3) Determining the Utopia Point: The utopia point, $\mathbf{a}$, represents a point in utility space desired by all users, and is defined in Equation 2. The values u_i^{max} described represent the maximum utilities for each user. Thus, in the proposed rate allocation system, each u_i^{max} value will be equal to the maximum utility possible for the given speech coder. In the case of the GSM AMR-NB encoder used, the maximum possible utility value can be obtained from Table I.

4) Determining Normalized Bargaining Powers With Equal Priorities: In the most common scenario, all users in the M-user proposed network will have equal priority. In other words, the speech quality of the transmitted signal from user i is of equal importance to the overall system as the quality of speech of the transmitted signal of user j, for $1 \leq i, j \leq M$. The following section derives the formula for bargaining powers in the equal priority case.

In the general KSBS, the bargaining powers of users can be interpreted as the relative weights of the user's demands. In the mathematical solution of the KSBS, the bargaining powers of the users in the system are given by the matrix C. As stated previously, C is a diagonal matrix with the c_{ii} element representing the normalized bargaining power of user i.

As proposed earlier, we use parameters of the input speech signal of user i to determine the corresponding normalized bar-

gaining power, which will be referred to as ρ_i. The equations for the bargaining powers of speech signal $\mathbf{s}(n)$ are given in Section III-D. We now define C as:

$$C = \text{diag}\{\mathbf{p}\}, \tag{13}$$

where

$$\mathbf{p} = [\rho_1, \rho_2, \ldots, \rho_M]^T, \tag{14}$$

diag $(\mathbf{p})$ is the matrix with diagonal values of $\rho_1, \ldots, \rho_M$ and zero-valued nondiagonal elements, and where the user normalized bargaining powers ρ_i are given by:

$$\rho_i = \frac{\Gamma_i}{\sum_{m=1}^{M} \Gamma_m}, \tag{15}$$

where Γ_i represents the unnormalized bargaining power of user i.

5) Determining Normalized Bargaining Powers With Weighted Priorities: In certain possible scenarios, the priorities of different users in the M-user network may differ from each other at given times. These scenarios might arise in emergency situations, for example, when a certain user is considered more important than the other users. Therefore, formula for weighted bargaining powers are necessary.

Let the vector $\mathbf{w} \in R^{M \times 1}$ contain the relative priority weights of the M users. That is:

$$\mathbf{w} = [w_1, w_2, \ldots, w_M]^T, \tag{16}$$

where w_i corresponds to the relative priority weight of user i. It then follows intuitively that $\mathbf{p}_w$, the vector of normalized weighted bargaining power, can be calculated as follows:

$$\mathbf{p}_w = [\rho_1^w, \rho_2^w, \ldots, \rho_M^w]^T, \tag{17}$$

where

$$\rho_i^w = \frac{w_i \Gamma_i}{\sum_{m=1}^{M} w_m \Gamma_m}. \tag{18}$$

Note that when the priority weight vector is set to $\mathbf{w} = [1, 1, \ldots, 1]$, the formula in Equations 17 and 18 simplify to the equal priority case of Section III-C.4.

6) Determining the KSBS in the Utility Space: With necessary bargaining parameters defined for our speech communication system, we can now apply our KSBS to the problem of rate allocation for multi-user speech transmission. Given the total rate constraint, we can define the quality feasibility set as discussed in Section III-B and then the Pareto Surface as discussed in Section III-C.1. Furthermore, we can position our disagreement point as discussed in Section III-C.2. Then, the Kalai-Smorodinsky bargaining solution can be stated simply as $\widehat{\mathbf{q}}$, where $\widehat{\mathbf{u}} = U(\widehat{\mathbf{q}})$, and where:

$$\widehat{\mathbf{u}} = \arg\max_{\mathbf{u}} \mathbf{p}^T \cdot \left(\frac{\mathbf{u} - \mathbf{d}}{|\mathbf{u} - \mathbf{d}|} \right), \tag{19}$$

for $\forall \mathbf{u} \in PS$.

Due to the discrete nature of the Pareto Surface, $\widehat{\mathbf{u}}$ represents an approximation to the determined KSBS. Thus, a certain amount of error between the calculated solution and $\widehat{\mathbf{u}}$ is unavoidable. However, this error is decreased for Pareto Surfaces with finer resolution, i.e. Pareto Surfaces consisting of a greater number of points. As discussed in Section III-C.1, finer resolution on the Pareto Surface is achieved with a speech encoder that can operate in a large number of quality modes.

7) Determining the KSBS in the Resource Space: The KSBS solution is determined in the utility domain as described in Section III-C.6. However, since the utility function $U(\cdot)$ is not 1-to-1, there does not exist an inverse function $U^{-1}(\cdot)$. Thus, in order to obtain a point in the resource domain, $\mathbf{q}_k$, such that:

$$\mathbf{u}_k = U(\mathbf{q}_k), \tag{20}$$

we rely on a predetermined codebook containing $(\mathbf{u}_k, \mathbf{q}_k)$ pairs to find our final rate allocation solution vector, $\widehat{\mathbf{q}}_k$. Our codebook is designed to not contain any mappings of distinct resource vectors, q_j and q_k, to the same utility vector, so that $u_l = U(q_j) = U(q_k)$. Determining a KSBS in the resource domain involves a search of the previously described codebook to match the utility domain KSBS, and since the codebook lists $(\mathbf{u}_k, \mathbf{q}_k)$ pairs, the resource domain KSBS can easily be obtained.

D. Defining Bargaining Powers Based on Short-Time Energy Level

Application of the KSBS to our problem of rate allocation necessitates the definition of bargaining powers. Bargaining powers represent the relative importance of additional rate to the users with respect to utility. This paper introduces a novel bargaining power based on the short-time energy of the input speech signal.

Scalable or multi-rate speech coders often include a hard decision Voice Activity Detection (VAD) algorithm [16] to classify speech versus non-speech segments. For example, the GSM AMR-NB speech coder includes the option to run at a lower bitrate when the current block of speech is determined to be comprised of solely background noise [13]. VAD algorithms often include a weighted sum of classifier functions, such as periodic similarity, zero crossing rate, spectral tilt, pre-emphasized energy ratio, and total frame energy. In such algorithms, the weighted sum of classifier functions is compared to a predetermined threshold. However, due to the possibly drastic effects of incorrectly classifying a speech segment as non-speech, the hard decision threshold is often set relatively low [16].

The proposed bargaining powers based on short-time energy offers a soft decision version of the VAD algorithms previously described. However, a linear function of the short-time energy will not serve as an efficient bargaining power since speech segments vary greatly in energy level, even though they generally contain more energy than non-speech segments. Therefore, we introduce a bargaining power based on a nonlinear function of the short-time energy which incorporates the A-Law companding function [18]. The A-Law function has been widely used as a quantization scheme in Pulse Code

Modulated (PCM) speech coding. We define the average short-time energy of the speech signal $s(n)$ as:

$$E_{ST} = \frac{\alpha}{N} \sum_{n=0}^{N-1} s^2(n), \qquad (21)$$

where N is the length of the speech signal segment, which was set to $N = 160$ in the proposed system, and α is an experimentally determined constant. The short-time energy bargaining power, Γ, is then defined as:

$$\Gamma = \frac{A \cdot E_{ST}}{1 + \log_{10}(A)} \text{ for } 0 < E_{ST} \leq \frac{1}{A} \qquad (22)$$

$$\Gamma = \frac{1 + \log_{10}(A \cdot E_{ST})}{1 + \log_{10}(A)} \text{ for } \frac{1}{A} < E_{ST} \leq 1 \qquad (23)$$

$$\Gamma = 1 \text{ for } E_{ST} > 1. \qquad (24)$$

In the proposed system $A = 68$, which is a common value in PCM algorithms, and α was found to give good performance at a value of $\alpha = 160$ when the input speech signals were normalized by the maximum value of the current utterance.

IV. COMPLEXITY AND SYSTEM LEARNING ISSUES

A. Complexity Analysis

The proposed rate allocation scheme uses a low complexity algorithm that can run in real-time. The KSBS algorithm for the M-user case involves the following computations at each block iteration:

- Calculating the bargaining power at each speech encoding block
- Normalizing the bargaining powers of each user at the CSM
- Determining the solution in the utility domain at the CSM
- Identifying the solution in the resource domain corresponding to the utility domain solution at the CSM

Calculating the bargaining power of a segment of speech for an individual user involves determining the short-time energy and the corresponded A-Law companded value. Let N_w represent the length of the window used in processing and coding the input speech signals. (In the GSM AMR-NB speech coder, $N_w = 160$.) The calculation of E_{ST} then involves N_w multiplications and $N_w - 1$ additions. The A-Law companding function involves 2 additions, 2 multiplications, and 2 logarithmic functions.

Determining the solution in the utility domain involves searching through a codebook to find the minimum result of a cost function. The calculation of the cost function, given in Equation 19, requires M subtractions, M multiplications, and 1 division. Let us define N_{PS} as the number of points on the Pareto Surface. Note that N_{PS} is a function of the number of users, M, and the total rate constraint, R. Additionally, once the solution in the utility domain has been located, the final solution in the resource domain can be easily looked up since the vectors are listed in pairs.

Table II summarizes the computational complexities of the operations within each speech encoder block, which can be

TABLE II

Computational Complexity of Operations Within Speech Encoder Blocks, where N_w Represents the Length of the Window Used for Coding in Samples

Task	Multiplications	Additions	Log functions
Determining BPs	$N_w + 2$	$N_w + 1$	2

TABLE III

Computational Complexity of Operations Within the CSM where M Represents the Number of Users and N_{PS} Represents the Number of Operating Points on the Pareto Surface

Task	Multiplications	Additions	Comparisons
Normalizing BPs	M	$M - 1$	0
Determining $\hat{u}_k$	$N_{PS} \cdot (M + 1)$	$N_{PS} \cdot M$	N_{PS}
Determining $\hat{q}_k$	0	0	1

carried out in parallel. Table III summarizes the computational complexities of the operations within the CSM.

As can be interpreted from Figure 2, the proposed system is integrated into the speech coding process. It is important to note that the computational load introduced by the rate allocation algorithm is far smaller than those introduced by other speech communication tasks such as encoding. Modern CELP speech coders include construction of both a short-term prediction filter and long-term prediction filter, approximation of pitch and pitch delay, and an extensive codebook search for optimum excitation positions and gains [16], and thus the proposed system has little effect on the computational load of the overall rate allocation and coding system.

Furthermore, our algorithm does not require any additional buffering other than the buffering necessary for block-based speech coding. In our implementation, the input speech was windowed and processed with a 20 ms window. Thus, our rate allocation system can be run in real-time.

B. Learning Process of the Proposed System

The proposed rate allocation system involves a low complexity training process. The only information necessary for the system is the matrix of rate-utility pairs, Θ, that represents the rate-performance curve of the chosen speech coder in terms of the chosen speech quality metric. The matrix Θ is defined as:

$$\Theta = \begin{bmatrix} u_1 \; r_1 \\ u_2 \; r_2 \\ \vdots \; \vdots \\ u_P \; r_P \end{bmatrix}, \qquad (25)$$

where u_i is the utility measure resulting from encoding speech at rate r_i, and P is the number of modes for the chosen multi-rate speech coder. The matrix Θ is estimated empirically by averaging the utility measures resulting from a given rate over a set of training speech segments. Let us define $u_{k,i}$ as the utility measure resulting from encoding the kth training speech

segment at a rate of r_i. The elements of the rate-performance matrix Θ can then be determined with the following equation:

$$u_i = \frac{1}{K} \sum_{k=0}^{K-1} u_{k,i}, \qquad (26)$$

where K is the number of speech segments used for training. This process is carried out for each mode, i.e. for $0 \leq i \leq P - 1$, of the chosen speech encoder to obtain the matrix Θ.

V. RESULTS

The proposed rate allocation system was tested on continuous speech segments from the DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus [19]. A subset of 275 sentences were randomly selected for training, 169 of which were spoken by various male speakers, and 106 of which were spoken by various female speakers. A subset of 255 sentences were randomly selected for testing, 168 of which were spoken by various male speakers, and 87 of which were spoken by various female speakers. Both the training set and testing set were spoken by the same group of speakers.

A. Baseline Allocation Systems

In order to show relative improvement of the proposed bargaining-based rate allocation algorithm, baseline algorithms must be defined. We introduce a basic scheme involving uniform allocation of a total rate of R to M users within the speech communication network, referred to as fairly allocated (FA) uniform allocation. In this algorithm, r_i, the amount of rate allocated to user i, is determined as:

$$r_i = \left\lfloor \frac{R}{M} \right\rfloor_{coder}, \qquad (27)$$

where the function $\lfloor \cdot \rfloor_{coder}$ returns the rate of the highest possible mode of the given speech coder whose bitrate is less than or equal to $\frac{R}{M}$. The FA uniform allocation scheme guarantees fair rate allocation among users, but proves to be very inefficient, as it generally cannot maximally allocate the given total rate R.

Therefore, a maximally allocated (MA) uniform allocation scheme is introduced. The MA uniform allocation algorithm involves the following steps:

- Determine initial user rates, $(r_1^o, r_2^o, \ldots, r_M^o)$ according to the uniform allocation scheme shown in Equation 27.
- Determine the excess rate, $\Delta R = R - M \cdot \lfloor \frac{R}{M} \rfloor_{coder}$.
- For users $i = 1, \ldots, M$, if ΔR is greater than or equal to the amount of rate needed to increase the encoding mode of user i, then increase the encoding mode of user i and update the excess rate.

B. Illustration of Example Input Signals to 2-User Rate Allocation System

The proposed rate algorithm was tested in the 2-user case, with equal priority weighting, and was tested using the proposed bargaining-based system, as well as with the

baseline algorithms of FA uniform allocation and MA uniform allocation. Randomly selected input speech signals, along with the corresponding bargaining powers, normalized bargaining powers, and allocated rates are shown in Figure 4. Panels 1 and 5 show the input speech waveforms to the 2-user system. Panels 2 and 6 illustrate the short-time energy bargaining powers corresponding to the input speech signals. Panels 3 and 7 show the resulting normalized bargag powers, and panels 4 and 8 show the rates allocated in time to users 1 and 2, respectively. As can be concluded from Figure 4, there is noticeable correlation between the presence of speech segments in the input signals and the corresponding short-time energy bargaining powers. Figure 4 also shows how the normalized bargaining powers translate into allocated rate.

Note the example time instant at $t = 0.90$ seconds, shown by the vertical dashed line. Within this speech block, the speech waveform of user 1 shows a high amplitude voiced signal, and the speech waveform of user 2 shows a low amplitude waveform. Note that the unnormalized and normalized bargaining powers become high for user 1 and become low for user 2. Finally, note that the bargaining powers translate into a high bitrate of $r_1 = 12.2$ kbps for user 1 and a low bitrate of $r_2 = 4.75$ kbps for user 2.

C. Increased Short-Time Coding Bitrate Resolution

The proposed rate allocation algorithm provides an increased short-time coding bitrate resolution. The short-time coding bitrate is defined in this paper as the mean rate for a user over a period of 2.0 seconds. In the case of the GSM AMR-NB speech codec used, the encoder can operate in 8 different bitrate modes, ranging from 4.75 kbps to 12.2 kbps, as shown in Table I. In the baseline scheme of MA uniform rate allocation, this results in poor short-time coding bitrate resolution. The proposed bargaining-based rate allocation algorithm provides improved short-time coding bitrate resolution relative to the baseline scheme of MA uniform allocation. Due to the dynamic nature of the rates allocated to each user in time, the proposed system can run at a large number of operating points. Figure 5 illustrates the bitrate operating points for the 2-user baseline network and the 2-user bargain-based network, for varying levels of total rate constraint. The utility used is the U_{SNR} utility function, discussed in Section III-B.

Note the greater number of operating modes available for the proposed rate allocation system. For example, in the 2-user case with a total rate constraint of 9.0 kbps per user, the only operating modes for the MA uniform allocation system are $(r_1 = 7.4, r_2 = 10.2)$, $(r_1 = 10.2, r_2 = 7.4)$, and $(r_1 = 7.4, r_2 = 7.4)$. The only operating mode for the FA uniform allocation system is $(r_1 = 7.4, r_2 = 7.4)$. However, as can be noted in Figure 5, there exist a large number of short-time operating modes for the bargaining-based system, and thus the proposed system can allocate rate to better match the characteristics of the input speech signals.

D. Rate Allocation for Continuous Speech

The proposed rate allocation system was then tested over 255 sentences. Table IV shows the results of the system on
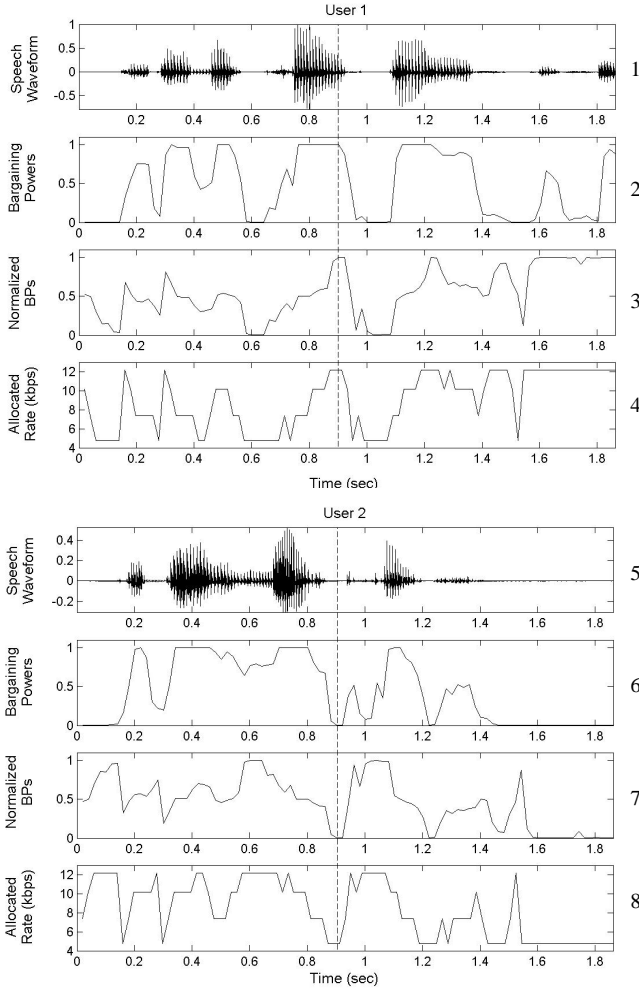
Fig. 4. *Example Input Speech Signals and Corresponding Bargaining Powers, Normalized Bargaining Powers, and Allocated Rates for the 2-user Case*
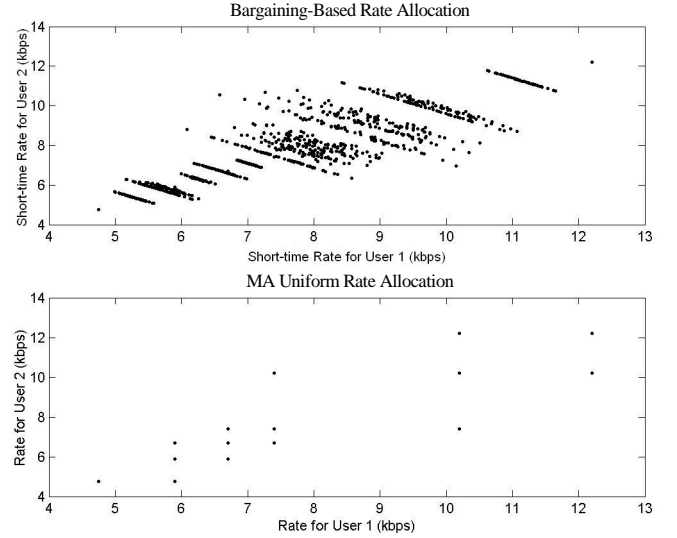

Fig. 5. *Operating Modes for the 2-User Baseline Network and the 2-User Bargaining-Based Network, for Total Rate Constraints Ranging from 9.5 kbps to 24.4 kbps: The top panel refers to bargaining-based rate allocation while the bottom panel refers to Maximally Allocated (MA) Uniform Allocation. Note that the operating points associated with Fairly Allocated (FA) Uniform Allocation are the subset of MA operating points along the diagonal axis.*

TABLE IV

Results for 2-User System Using the U_{SNR} Utility Function, Tested On Continuous Speech

Allocation Algorithm	$\overline{u}$	improvement over FA
FA Uniform Allocation	2.770 dB	N/A
MA Uniform Allocation	3.092 dB	0.322 dB
Bargaining-based Allocation	3.235 dB	0.465 dB

continuous speech signals, using the segmental SNR utility function. Table V shows the results of the proposed system using the Itakura utility function, as well as the I_{Itak} distortion function. Note that $\overline{u}$ represents the average utility of the input speech signals, and is given by:

$$\overline{u} = \frac{1}{M} \sum_{m=0}^{M-1} u_m, \tag{28}$$

where u_i represents the utility measure of user i, and M is the number of users in the system.

As can be concluded from Table IV and Table V the proposed rate allocation system provides increased utility measures relative to the baseline systems. The proposed algorithm provides 0.465 dB improvement over the FA uniform allocation scheme for the U_{SNR} utility value. Additionally, the proposed system provides a 12.84% decrease of the I_{Itak} distortion function. Additionally, the bargaining-based system guarantees fairness among users, which the MA uniform allocation system does not.

E. Rate Allocation for Simulated Conversational Speech

Since speech communication networks commonly transmit signals that are comprised of both speech segments and non-speech (silent) segments, the proposed system was tested on simulated conversational speech. The simulated conversational speech signals were created by concatenating randomly chosen sentences with silence of duration S_β, where S_β is a random variable with uniform distribution over the range $[0, \beta]$. Thus the expected value of S_β is $E[S_\beta] = \beta/2$. Figure 6 shows and example of simulated conversational speech signals for a 2-user network.

Table VI shows the results of the proposed rate allocation algorithm using the segmental SNR utility function, tested on

TABLE V

Results for 2-User System Using the U_{Itak} Utility Function, Tested On Continuous Speech

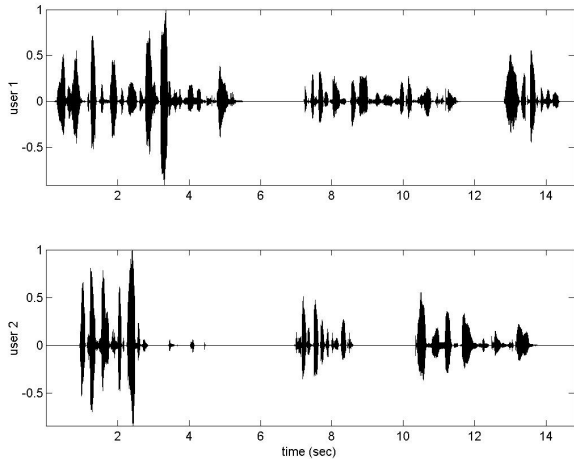| Allocation Algorithm	$\overline{u}$	I_{Itak}	% decrease from FA
FA Uniform Allocation	3.380	0.296	N/A
MA Uniform Allocation	3.813	0.262	−11.49%
Bargaining-based Allocation	3.880	0.258	−12.84%

Fig. 6. *Example Input Simulated Conversational Speech Signals for 2-user Case*

TABLE VI

Results for 2-User System Using the U_{SNR} Utility Function, Tested On Simulated Conversational Speech

Allocation Algorithm	$\overline{u}$	improvement over FA
FA Uniform Allocation	3.886 dB	N/A
MA Uniform Allocation	4.164 dB	0.278 dB
Bargaining-based Allocation	4.590 dB	0.704 dB

simulated conversational speech with $E[S_\beta] = 3.0$ seconds. Table VII shows the results of the proposed rate allocation algorithm using the Itakura utility function and the I_{Itak} distortion measure, also with $E[S_\beta] = 3.0$ seconds.

It can be interpreted from Table VI and Table VII that the bargaining-based rate allocation algorithm results in increased utility measures relative to the baseline systems. The proposed system provides 0.704 dB improvement relative to the FA uniform allocation scheme in terms of the U_{SNR} utility functions. Additionally, the proposed system provides a 19.50% decrease in the I_{Itak} distortion measure. Furthermore, the relative improvements shown for simulated conversational speech are greater than those provided for continuous speech. Improved performance is due to the fact that users are typically not speaking simultaneously. For example, in the M-user example, if user i is silent for a period of time, all excess rate can be allocated to the other users in the system.

TABLE VII

Results for 2-User System Using the U_{Itak} Utility Function, Tested On Simulated Conversational Speech

Allocation Algorithm	$\overline{u}$	I_{Itak}	% decrease from FA
FA Uniform Allocation	3.146	0.318	N/A
MA Uniform Allocation	3.531	0.283	-11.01%
Bargaining-based Allocation	3.906	0.256	-19.50%

TABLE VIII

Performance of Weighted Priority Rate Allocation in the 2-User Case, Using the U_{SNR} Utility Function

Relative Weight Vector	$[1,1]^T$	$[1,2]^T$	$[1,3]^T$
r_1 (kbps)	8.749	7.960	7.807
r_2 (kbps)	8.851	9.640	9.793
u_1	3.136 dB	2.905 dB	2.869 dB
u_2	3.501 dB	3.744 dB	3.886 dB

F. Weighted Priority Rate Allocation

As discussed in Section III-C.5, there may exist scenarios in which the utilities of users may be weighted unequally. These relative weights are defined in vector form in Equation 16. Table VIII shows examples of the resulting rates and utilities for the 3-user case, for arbitrary relative weight vectors, using the U_{SNR} utility function.

In can be concluded from Table VIII that the resulting user rates and utility measures reflect the corresponding priority weighting vectors.

Several benefits are shown for the proposed bargaining-based rate allocation system for multi-user speech networks. Firstly, the proposed system results in superior quality of speech measures relative to the baseline system of uniform allocation. The second major benefit of the proposed algorithm is increased short-time coding bitrate resolution. Finally, the proposed system is robust to scenarios in which the quality of speech measures of users are weighted differently.

VI. CONCLUSION

This paper focuses on a non-collaborative multi-user speech communication system with a central spectral moderator. Specifically, this paper investigates bargaining theory as a method of rate allocation in an M-user system, and applies the generalized KSBS to solve this problem. The algorithm developed uses the concept of bargaining powers based on the short-time energy of input speech signals, and accordingly allocates rate. The rate allocation scheme is designed to be applicable to any multi-rate speech coder, and is robust to a variety of speech quality metrics.

The proposed rate allocation system is shown to provide increased speech utility measures relative to the uniform allocation baseline systems. Additionally, the system is shown to involve a quick and low complexity training process. It is also shown to be robust to scenarios in which the quality of speech of users are weighted differently. Finally, the proposed system is shown to improve the short-time coding bitrate resolution of the GSM AMR-NB speech coder. This results in a greater number of operational modes for the multi-user network.

VII. ACKNOWLEDGMENTS

REFERENCES

[1] P. Leaves, S. Ghaheri-Niri, R. Tafazolli, L. Christodoulides, T. Sammut, W. Stahl, and J. Huschke, *Dynamic Spectrum Allocation in a Multi-Radio Environment: Concept and Algorithm*, 3G Mobile Communication Technologies, 2001.

[2] P. Leaves, J. Huschke, and R.Tafazolli, *A Summary of Dynamic Spectrum Allocation Results from DRiVE*, IST Mobile Summit, 2002.

[3] A. Parekh and R. Gallager, *A generalized processor sharing approach to flow control in integrated services networks: the single node case*, IEEE/ACM Trans. Networking, vol. 1, no. 3, pp. 344357, June 1993.

[4] I. Forkel, T. Kriengchaiyapruk, B. Wegman, and E. Schulz, *Dynamic Channel Allocation in UMTS Terrestrial Radio Access TDD Systems*, Proc. IEEE Vehicular Technology Conference, 2001.

[5] F. P. Kelly, A. K. Maulloo, and D. K. H. Tan, *Rate Control for Communication Networks: Shadow Prices, Proportional Fairness and Stability*, Journal of the Operational Research Society, vol. 49, no. 3, pp. 237252, Mar. 1998

[6] S. Gunturi, and F. Paganini, *Game Theoretic Approach to Power Control in Cellular CDMA*, IEEE Vehicular Technology Conference, Orlando, vol. 4, 2003, pp 2362-2366

[7] J. Sun, E. Modiano, and L. Zheng, *Wireless Channel Allocation Using an Auction Algorithm*, IEEE Journal on Selected Areas in Communications, vol. 24, no. 5, May 2006.

[8] H. Park, M. van der Schaar, *Utility-based Fairness for Multi-user Wireless Multimedia Resource Allocation using Bargaining*, submitted to IEEE Journal on Selected Areas in Communications, special issue on Non-Cooperative Behavior in Networking, 2006.

[9] J. F. Nash, *The Bargaining Problem*, Econometrica 18, 155-162, 1950

[10] H. Raiffa, *Arbitration Schemes for Generalized Two-Person Games*, University of Michigan, 1951.

[11] C. A. Mattson, A. Messac, *Pareto Frontier Based Concept Selection Under Uncertainty, with Visualization*, Optimization and Engineering, Special Issue on Multidisciplinary Design Optimization, in press

[12] E. Kalai, M. Smorodinsky, *Other solutions to Nash's bargaining problem.*, Econometrica 43, 513-518, 1975

[13] ETSI, *Adaptive Multi-Rate Speech Codec*, Technical Report TS 126.090 v3, European Telecommunications Standards Institute, 2000

[14] M. Hasegawa-Johnson, A. Alwan, *Speech Coding: Fundamentals and Applications*, Wiley Encyclopedia of Telecommunications, 2003, John Wiley and Sons, Inc.

[15] L. Rabiner, B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993

[16] A. M. Kunduz, *Digital Speech*, John Wiley and Sons, Inc., 2004

[17] T. F. Quatieri, *Discrete-Time Speech Signal Processing*, Prentice Hall, 2001

[18] K. W. Cattermole, *Principles of Pulse Code Modulation*, Illiffe, 1973

[19] J. S. Garofolo,L. F. Lamel,W. M. Fisher,J. G. Fiscus,D. S. Pallett,and N. L. Dahlgren, *DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM*, NIST, 1993.