Introduction

Digital hearing aids have gained rapidly in popularity in recent years, now comprising over 93% of the US market (figure 1). Interestingly, despite this increase, customer satisfaction, benefit, use, and residual disability have not changed appreciably during the past decade (figure 2). One hearing aid feature, however, that has produced considerable increases in patient satisfaction is the directional microphone (figure 3). Directional microphones, of course, do not require the use of digital technology, as they have been available commercially in hearing aids for over thirty years. That said, a recent trend has been to incorporate directional microphones into digital hearing aid systems that monitor the listening environment and automatically activate fixed- or adaptive-directional microphone arrays only when appropriate. These systems represent the first approximation of “acoustic scene analysis”, which involves a classification and decision-making process that can recognize a wide variety of sound environments and adapt the hearing aid characteristics accordingly.

Numerous studies have been published that suggest that automatic activation of hearing aid features may provide increased use by patients than when they are manually activated (Surr, Walden, Cord and Olson 2002; Cord, Surr, Walden and Dyrlund 2004). In addition to directional microphones, acoustic scene analysis may be used to adapt features including feedback reduction, noise/reverberation cancellation, and multiple-channel compression. The classification accuracy of this system has proven to be highly sensitive and specific in laboratory settings under simulated conditions (Buchler 2002). As these systems move from laboratory to clinic, however, it raises the issue of how to best evaluate classifier performance for “real-world” conditions. Further complicating this issue is the prevalence of “datalogging” features with modern digital hearing aids, which may be used to further optimize hearing aid performance to meet individual patient needs, moving from “acoustic scene analysis” to “auditory scene analysis”. The appropriate balance between controlled laboratory
Figure 2. Customer satisfaction with value, benefit in noise, overall benefit and likelihood of repurchasing current brand of hearing instrument (hearing instruments <3 years of age; source MarkeTrak III [1991] – MarkeTrak VI [2000]).

Figure 3. Factors showing at least 10% improvement in customer satisfaction (from Kochkin MarkeTrak VI 2004).
conditions and actual hearing aid use by individual patients challenges both clinicians and researchers. The present studies evaluated the performance of independent bilateral hearing aids in various listening environments to assess:

- The sensitivity and specificity of program switching behavior;
- Individual preference differences reported by patients with wearable devices;
- Symmetric versus asymmetric classification between ears.

Sound Classification System Overview

The basic building blocks of any sound classification system are depicted in figure 4.

![Sound Classification System](image)

**Figure 4.** Block diagram of a typical acoustic scene analysis system.

**Sounds**

The development of a library of relevant sounds to an average hearing aid user is pre-requisite to the development of any acoustic scene analysis system. In their 1991 study, Fedtke, Fuder, Hamann and Haubold asked subjects with moderate hearing loss to judge the importance of 52 different listening situations related to home life, work, culture, leisure and transportation. The situations judged most important were segregated into four main categories:

- Speech (dialogue, theatre, cinema, phone, TV);
- Speech in noise (cocktail party, announcements at train station or airport, speech in car);
- Alarm signals (ringing phone, door bell, smoke detector);
- Nature (chirping birds).

More recently, Buchler (2002) classified acoustic signals into five main acoustic situations, which included speech in quiet and noise, plus noise alone (including warning and alarm signals), music, and interestingly, silence. He also developed a library of nearly 300 sound samples for speech in a variety of different listening environments, plus “noise only” samples, and music from a variety of genres to define an appropriate library of sounds for evaluation by the prototype classification system.

**Feature Extraction**

Previous researchers have used a variety of feature extraction techniques for feature classification of different acoustic environments (e.g. Ludvigsen, 1993, Kates, 1995). The auditory mechanisms evaluated may include spectral and temporal analysis along several dimensions, including:

- Overall amplitude and level fluctuations;
- Temporal fluctuations – onset/offset patterns;
- Spectral profile – spectral “center of gravity” and fluctuations;
- Harmonicity – tonality and pitch variations;
- Input signal type – acoustic, inductive, or FM;
- Inter-aural time and level differences;
- Amplitude modulation;
- Frequency modulation.

Independently, each of these features may provide some differences between the respective sound classes. For example, the feature “tonality” may be used to distinguish between “music” and “noise” with little overlap, but not between “speech in quiet” and “music” (figure 5, based on Buchler 2002). As a result, several rule-based and statistical classification systems were developed, using different feature combinations, to arrive at those that provided the highest sensitivity and specificity for classification of “library” of relevant acoustic sounds listed above into the four sound categories of Quiet, Speech in Noise, Noise Alone, and Music.

**Classification and Class Decision**

The accuracy of any feature classification system is dependent on how well the feature parameters differentiate between sound categories. An example of this is depicted in figure 6, which illustrates a “feature space” defined for a classifier designed to choose between four sound classes (speech in quiet, speech in noise, noise
alone, and music). The degree of overlap between sound categories is inversely proportional to the accuracy of the sound classifier.

Buchler (2002) evaluated 19 acoustic features and six classification systems, using the sample acoustic library and the four acoustic situations (speech in quiet, speech in noise, noise only, and music). Confusion matrices were developed (see figure 7) for each classification system for repeated presentations to the prototype hearing instruments, and receiver operating characteristics (ROC) indicated the sensitivity and specificity of each classification system. Initial results indicated that the best classification performance was provided with an ergonomic Hidden Markov Model (HMM) with the feature set comprising pitch, amplitude modulation, spectral and onset features (figure 8). These results were improved further when combined with a multi-stage classifier system that improved hit rates to over 92%, with false alarm rates less than 3%, for three of the four acoustic situations. The remaining class (speech in noise) has a hit and false alarm rate of 87% and 5%, respectively. These findings indicate that acoustic scene analysis may be used in wearable devices to provide very accurate classification of multiple acoustic scenes in wearable hearing aids. Of additional interest, however, was the reaction of human subjects when fit with wearable devices.

**Figure 5.** Mean feature value, tonality (from Buchler 2002).

**Figure 6.** Illustration of how two feature parameters may be used to characterize a feature space that differentiates between four sound categories (speech, noise, speech in noise, and music).
A more sophisticated version of this prototype system was evaluated for use in hard-of-hearing patients under "real world" situations. The acoustic classification system (AutoPilot) monitors and switches hearing aid programs automatically between four acoustic scenes:

- Calm (Quiet situations);
- Speech in Noise;
- Comfort in noise;
- Music.

This use of acoustic scene analysis for program switching was monitored in 180 hard-of-hearing patients who were fit binaurally and wore the instruments for a period of four to six weeks. Datalogging was used to automatically monitor several dimensions of hearing aid use, including:

- Overall usage of the hearing instrument, in hours and average hours per day;
- Automatic and manual program selection, in percent of total usage;
- Usage of the different programs in manual program mode;
- Averaged volume control adjustments for each automatic base program.

In addition, patients were asked rate the accuracy, frequency, and speed of program switching between the four acoustic environments. A self-assessment questionnaire was used to express preferences for AutoPilot acoustic scene analysis settings using a five-point scale.

### Confusion Matrix

<table>
<thead>
<tr>
<th>Original</th>
<th>Speech</th>
<th>Speech in noise</th>
<th>Noise</th>
<th>Music</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech</td>
<td>4524</td>
<td>400</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>Speech in noise</td>
<td>374</td>
<td>4931</td>
<td>404</td>
<td>291</td>
</tr>
<tr>
<td>Noise</td>
<td>0</td>
<td>524</td>
<td>5728</td>
<td>248</td>
</tr>
<tr>
<td>Music</td>
<td>191</td>
<td>588</td>
<td>135</td>
<td>5086</td>
</tr>
</tbody>
</table>

Figure 7. Sample confusion matrix indicating original category (ordinate) and classification (abscissa).

**Datalogging**

A more sophisticated version of this prototype system was evaluated for use in hard-of-hearing patients under "real world" situations. The acoustic classification system (AutoPilot) monitors and switches hearing aid programs automatically between four acoustic scenes:

- Calm (Quiet situations);
- Speech in Noise;
- Comfort in noise;
- Music.

### ROC for Hidden Markov Model

![ROC for Hidden Markov Model](image)

Figure 8. Receiver Operating Characteristic (ROC) for Hidden Markov Model (HMM) Classification System indicating the best combination of hit rate and low false alarm rates that were observed for the feature set that used tonality, width, center of gravity, and onset cues (from Buchler 2002).
Figure 9. Software used to fine-tune acoustic scene analysis parameters to optimize for individual patient preference.

Figure 10. Subject responses to the question, “Is the AutoPilot adaptation appropriate?” after two-week trial period. Based on multi-national study of 180 subjects.
The results indicated uniformly positive sound quality and speech understanding in both quiet and noisy listening situations. Software was used to fine-tune the individual base programs when subjects indicated the need for further optimization (figure 9). Initially, according to the DataLogging feature, the average subject used the “AutoPilot” program 67% of the time, and noticed appropriate switching between the four acoustic scenes. After the initial follow-up fitting, 75% reported that the AutoPilot adaptation was “often” or “mostly” appropriate (figure 10). When software was used to further fine tune these settings in response to individual preference, this percentage increased to nearly 90% by the second follow-up fitting. These findings suggest that although acoustic scene analysis may be developed for use with “automatic” hearing instruments, personal patient preference may be used to further compensate these settings to meet individual patient preferences.

Asymmetric Switching

One possible reason for the large inter-subject differences in figure 10 may be due to large differences in the “real-world” environments encountered by different hearing aid users. For example, some users may spend more time in low-ambient noise environments, while others may be in symmetric, asymmetric (left/right) or diffuse listening environments. Analysis of datalogging results revealed no statistically-significant relationship between “wrong” program selection and actual hearing aid use in the four different “AutoPilot” programs (calm, speech in noise, comfort in noise, or music). Consequently, the issue of asymmetric listening environments and program switching behavior was explored.

A single human subject was fitted with independent ear-level bilateral hearing instruments that served as recording microphones for different “real-world” listening environments. Independent outputs from each device served as output into a central processor for analysis. A total of 22 hours of various listening hours were recorded, comprising a total of 63 separate “tracks” that ranged from two minutes to two hours in duration. For each track, the output of the device was constrained to one of the four acoustic classes used in AutoPilot: speech in quiet, speech in noise, noise alone, or music. The central processor was used to analyze synchronous and asynchronous switching behavior over the 22 hours recording time.

Results are illustrated in figure 11, and indicate that across the 63 “real world” tracks, symmetric program switching (output from both ears the same) was present 93% of the time. For this test condition, the speech in quiet program was classified 16% of the time in both devices, with speech in noise, noise alone, and music classified at 31%, 39%, and 7%, respectively. This ratio of quiet to noisy listening backgrounds confirmed previous studies (e.g., Walden, Surr. Cord and Dyrlund 2004) that indicated directional microphones are preferred by hearing aid users approximately 33% of the time for typical hearing aid users. In the Walden et al. study, subjects

- 93% symmetric situations:
  —39% Noise/Noise
  —31% Speech in Noise/Speech in Noise
  —16% Speech in Quiet/Speech in Quiet
  —7% Music/Music
- 7% asymmetric situations

Figure 11. Results of bilateral classification output for 63 real-world listening environments.
also preferred omni-directional microphone settings 37% of the time and had no preference 30% of the time.

For the 7% of the time when outputs from the left and right instruments were asymmetric, the vast majority of occurrences were when noise was present in the ambient listening environment (figure 12). For 47% of these occurrences, speech in quiet was detected in one ear, indicating that the primary talker may have been proximal to one ear. Although objective speech recognition evaluation was not conducted, this may provide an advantage over binaural synchronization. Additional experimentation is underway to explore this possibility.

The mean length, in seconds, of program asymmetry is indicated in figure 13. The median length of program asymmetry, when present, was 1.8 seconds. Very few asymmetries in excess of 20 seconds existed, and were restricted to slightly noisy to rather noisy (60 dB SPL or greater) situations combined with a conversation with one or a few individuals situated to one side of the user (e.g. whispering in the movie theater, chatting while sitting at a bar, discussion when seated next to another person in an auditorium, and talking while cooking on a stove). These data reinforce that it is beneficial to allow independent switching behavior to optimize communication in these listening situations. Preliminary data suggest that a more important factor may be to focus on monitoring and adjusting for bilateral, situation-specific volume control adjustments made by hearing aid users across listening environments. Additional research is required, however, to confirm these findings. The bottom line is that hearing instrument technology continues to become more and more sophisticated; the challenge becomes combining laboratory data with real-world experience to find the most appropriate evaluation strategies for predicting patient benefits.

Conclusions

A four-scene acoustic classifier was developed with very high sensitivity and specificity under controlled laboratory situations that simulated “real world” listening environments. Hearing-impaired subjects who wore devices in-situ report range of preference judgments for switching behavior, which suggests the need for individual optimization. Evaluation of symmetric and asymmetric switching behavior, however, illustrates the
complexity of design and evaluation of acoustic classifiers for laboratory and real-world use. In moving from “acoustic” to “auditory” scene analysis, it is important to consider bilateral ear independence as a critical component (among others) of scene classification.

References