

# Data Mining of Audiology Patient Records: Factors Influencing the Choice of Hearing Aid Type

Muhammad Naveed Anwar

PhD Researcher

Department of Computing, Engineering & Technology  
University of Sunderland, Sunderland  
0044-7508702055

naveed.anwar@sunderland.ac.uk

Michael Philip Oakes

Senior Lecturer

Department of Computing, Engineering & Technology  
University of Sunderland, Sunderland 0044-  
1915153631

michael.oakes@sunderland.ac.uk

## ABSTRACT

In this paper we describe our analysis of a database of over 180,000 patient records, collected from over 23,000 patients, by the hearing aid clinic at James Cook University Hospital in Middlesbrough, UK. These records consist of audiograms (graphs of the faintest sounds audible to the patient at six different pitches), categorical data (such as age, gender, diagnosis and hearing aid type) and brief free text notes made by the technicians. We mine this data to determine which factors contribute to the decision to fit a BTE (worn behind the ear) hearing aid as opposed to an ITE (worn in the ear) hearing aid. From PCA (principal component analysis) we determined four main audiogram types, and we relate these to the type of hearing aid chosen. We combine the effects of age, gender, diagnosis, masker, mould and individual audiogram frequencies into a single model by means of logistic regression. We also discovered some significant keywords in the free text fields by using the chi-squared ( $\chi^2$ ) test, which can also be used in the model. The final model can act a decision support tool to help decide whether an individual patient should be offered a BTE or an ITE hearing aid.

## Categories and Subject Descriptors

Primary Classification: H. [Information Systems]: H.2 DATABASE MANAGEMENT – H.2.8 Database Applications – Subjects: Data mining. Additional Classification: J. [Computer Applications]: J.3 LIFE AND MEDICAL SCIENCES – Subjects: Medical information systems.

## General Terms

Algorithms, Experimentation.

## Keywords

Chi-Squared, Hearing Aids, Logistic Regression, Principal Component Analysis.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*DTMBIO'11*, October 24, 2011, Glasgow, Scotland, UK.

Copyright 2011 ACM 978-1-4503-0960-8/11/10...\$10.00.

## 1. INTRODUCTION

In this paper we look for factors influencing the choice between two common hearing aid types: BTE (worn behind the ear) or ITE (worn in the ear). This choice is typically made by audiology technicians working in out-patient clinics, on the basis of audiogram results and consultation with the patient. In many cases, the choice is clear cut, but at other times the technicians might benefit from a second opinion given by an automatic system with an explanation of how that second opinion was arrived at. The production of such a decision support system is the main goal of this paper. Our data set is unusual in that ITE hearing aids are not generally available on the British National Health Service in England, as they are more expensive than BTE hearing aids. However, both types of aid are prescribed at James Cook University Hospital in Middlesbrough, UK. The data, collected between 1992 and 2001, consists of the following types of records:

1. Audiograms (graphs of the auditory thresholds, or faintest sounds audible to the patient at six different pitches or frequencies, where 0 shows perfect hearing and higher thresholds show impaired hearing), e.g., 40, 35, 35, 35, 85, 70, 15, 20, 20, 30, 55, where the first six values are AC (air conduction) and the last five are for BC (bone conduction). AC is measured by placing headphones over the ears, and determines the overall level of hearing. BC is measured by placing the sound source tightly on the mastoid bone behind the ear, and measures the level of hearing of the inner part of the ear. A constraint on the data is that BC must always be the same or better than AC. The difference between the AC and the BC is called the air-bone gap, and measures the hearing ability of the middle and outer parts of the ear.
2. Categorical data (such as gender, diagnosis and hearing aid type), e.g., M, TINNITUS, BE18.
3. Brief free text notes made by the technicians, e.g., IMPS. TAKEN FOR BINAURAL AIDS., where IMPS is an abbreviation for “impressions”, and BINAURAL means “worn in both ears”.

## 2. PRINCIPAL COMPONENT ANALYSIS ON AUDIOGRAMS

In this section, we describe how we performed a Principal Component Analysis (PCA) on the set of 11,462 audiograms where all AC and BC thresholds for the right ear were recorded, to determine the main audiogram types found among hearing aid

users. We then assigned each patient to one of these main audiogram groups, and used the chi-squared test to look for any association between hearing aid type and audiogram type. The rows of our input matrix were the individual audiograms for the right ear, while the 11 columns were for six air conduction and five bone conduction thresholds. Although the patients were originally tested at 11 frequencies, the principle of PCA is that certain frequencies tend to vary together, and thus can be grouped into a smaller number of underlying variables called principal components (PC). Each PC has a set of coefficients in the range -1 to +1, corresponding to the degree of influence of each of the original thresholds on that PC.

**Table 1. Component coefficient vectors of PCA**

	PC1	PC2	PC3	PC4
AC250	-0.3001	-0.3811	0.2988	-0.1677
AC500	-0.3218	-0.3619	0.2754	-0.0166
AC1000	-0.3410	-0.1999	0.2427	0.2643
AC2000	-0.3436	0.1440	0.1910	0.2697
AC4000	-0.3031	0.3673	0.2409	-0.1742
AC8000	-0.2722	0.3186	0.2629	-0.4684
BC250	-0.2510	-0.2304	-0.4890	-0.5087
BC500	-0.2942	-0.2404	-0.4152	-0.0846
BC1000	-0.3189	-0.0760	-0.3052	0.3595
BC2000	-0.3028	0.2699	-0.2419	0.4088
BC4000	-0.2516	0.4870	-0.2219	-0.1299

The coefficients of the first PC (PC1) were all negative and approximately equal. This suggests that the main source of variation between the patients was simply the overall degree of hearing loss. The coefficients of the second PC (PC2) were negative for frequencies at or below 1000Hz, but positive for higher frequencies, for both air and bone conduction, and thus differentiated patients according to whether they have a predominantly high frequency or low frequency hearing loss. The coefficients of the third PC (PC3) were positive for air conduction at all frequencies, but negative for bone conduction, showing a contrast between patients with and without an air-bone gap. The fourth component (PC4) is similar to the third, but corresponds to an air-bone gap at low frequencies. No clear patterns were seen for the fifth or subsequent principal components. The first four PCs correspond to audiogram types frequently encountered in audiology clinics. The percentage of the overall variability in the data explained by the first four principal components respectively was 59.5, 13.4, 9.7, and 5.2, giving a total of 87.8%.

Our method of converting PCA scores to typical audiogram types is an approximation only. It assumes that the PCA scores are directly related to audiogram thresholds, whereas in reality they measure different things: PCA measures the importance of a threshold in distinguishing audiograms, while audiograms measure the degree of hearing loss at the same frequency. The range of PCA scores is -1 to +1, while the range of audiogram thresholds is 0 to 120 dB. If we assume that the relationship is linear, then for each frequency a PCA value of -1 corresponds to an audiogram threshold of 0, and a PCA value of 0 = an audiogram threshold of 60, and a PCA value 1 = an audiogram threshold of 120. The formula relating the two was:

$$\text{Audiogram\_threshold} = 60 + (60 \times \text{PCA\_score}). \quad (1)$$

The thresholds corresponding to the first four PCs by using (1) are given below. The top line refers to the air conduction frequencies at 250, 500, 1000, 2000, 4000 and 8000Hz respectively, while the second line refers to the corresponding bone conduction frequencies in the range 250 to 4000Hz.

1. PCA 1 : Flat hearing loss

42	41	40	39	42	44
45	45	42	42	45	

2. PCA 2 : Pure high tone sensorineural (inner ear) loss

37	38	48	69	82	79
46	46	55	76	89	

3. PCA 3 : Air-bone gap (flat)

78	77	75	71	75	76
31	35	42	45	47	

4. PCA 4 : Air-bone gap (predominant at low tone)

50	59	76	76	50	32
29	55	72	75	52	

We classified each individual patient audiogram into one of the main audiogram types identified above, according to least Euclidean distance. A chi-squared test was then performed to determine whether there was any association between the audiogram class of each patient and the type of hearing aid worn. This test was done on the set of 7,437 records where all AC and BC thresholds were available for the right ear, and either a BTE or an ITE aid was specified. The results are shown in Tables 2 to 4.

**Table 2. Observed values (O)**

Hearing aid type	PCA1	PCA2	PCA3	PCA4
ITE	2036	1341	476	75
BTE	1119	1166	1165	59

**Table 3. Expected values (E)**

Hearing aid type	PCA1	PCA2	PCA3	PCA4
ITE	1666.38	1324.12	866.73	70.77
BTE	1488.62	1182.88	774.27	63.23

**Table 4.  $(O - E)^2 / E$  values**

Hearing aid type	PCA1	PCA2	PCA3	PCA4
ITE	81.99	0.22	176.14	0.25
BTE	91.78	0.24	197.18	0.28

The overall  $\chi^2$  (chi-squared value), calculated as the sum of the cells in Table 4, is 548.07, which for one degree of freedom gives  $p < 0.001$ , so audiogram type is clearly related to hearing aid type. Also, in Table 4, the  $(O - E)^2 / E$  values which make the greatest contribution to the overall  $\chi^2$  value are those in the PCA1 and PCA3 columns. Thus, flat hearing loss (PCA1) audiograms without air-bone were associated with ITE aids and flat

audiograms with additional air-bone gaps (PCA3) were associated with BTE aids. This result is in accordance with Stephens [1], who found that the fitting of ITE aids was limited in cases of severe hearing loss. We also clustered audiograms using K-means clustering [2] and found similar results to PCA, in that the mild to moderate hearing loss cluster was associated with ITE aids and the severe hearing loss cluster was associated with BTE aids. We thus demonstrated that the audiogram is a factor influencing the choice of hearing aid type. In our final logistic regression model, rather than simply using the four identified broad audiogram types, we used each individual hearing threshold.

### 3. USE OF THE CHI-SQUARED TEST TO DISCOVER OTHER FACTORS RELATED TO HEARING AID TYPE

In the previous section we showed how we found that the choice of hearing aid type was related to the shape of the audiogram. In this section we will describe how we used the simple chi-squared test to discover which of the category data fields were significantly associated with the choice of hearing aid type, and also to discover free-text keywords which were significantly associated with either BTE or ITE hearing aids.

The contingency table showing the relationship between gender and hearing aid type is shown in Table 5. The raw counts are given at the top of each cell, where for example there were 3196 male patients who wore BTE hearing aids. In each cell the Observed frequencies ( $O$ ) are not enclosed in brackets, Expected frequencies ( $E$ ) are in ( ) and the quantity  $(O - E)^2 / E$  is in [ ]. The overall chi-squared value (the sum of the values in [ ] for all four cells) was 33.68, which for one degree of freedom is significant at  $p < 0.001$ . Males tended more to use ITE hearing aids and females tended more to use BTE hearing aids. For the relationship between hearing aid type and a diagnosis of tinnitus (ringing in the ear), the overall chi-squared value was 31.75, again significant at  $p < 0.001$  for one degree of freedom. Patients with tinnitus tended more to wear ITE hearing aids. The relationship between the wearing of a tinnitus masker (a soothing sound source designed to drown out tinnitus) and hearing aid type, among patients diagnosed with tinnitus, had the overall chi-squared value was 17.16, which for one degree of freedom, was also significant at  $p < 0.001$ . Our data for the cross-tabulation of hearing aid type and age had the overall chi-squared value was 10.53, which for one degree of freedom, showed significance at  $p < 0.001$ . Mould type is also cross-tabulated with hearing aid type and the overall chi-squared value was 9844.18, which for 30 degrees of freedom was significant at  $p < 0.001$ . Thus all our category data types were significantly associated with hearing aid type. We used all the data in the patient records without considering confounding effects, where for example it might have been the choice of hearing aid type affecting the choice of mould, rather than vice versa. We believe that this may have been the case, since many mould types never occurred in conjunction with one or the other hearing aid type.

To discover those free-text keywords which were significantly associated with either BTE or ITE hearing aids, first a large contingency table was created where the rows stood for hearing aid type, while each column stood for a candidate keyword (one of 664 distinct words found to have occurred at least once in the free-text fields of the entire record set). The observed value in each cell was the number of times that word had been found in the free-text fields of patients with that type of hearing aid. This table

had an overall chi-squared value of 5421.84 for 663 degrees of freedom, giving  $p < 0.001$ . This data show, with 99.9% confidence that these free text words were not randomly distributed, but some text words are associated with hearing aid type.

**Table 5. Observed and expected frequencies for ITE/BTE aid with gender**

Hearing aid type	Male	Female	Row total
BTE	3196 (3369.38) [8.92]	3850 (3676.62) [8.17]	7046
ITE	3647 (3473.62) [8.66]	3617 (3790.38) [7.93]	7264
Column total	6843	7467	14310

To find the association between individual free-text words and hearing aid type, we examined the quantity  $(O - E)^2 / E$  to rank the keywords according to importance. The set of free-text keywords which tended to occur significantly more and less often (called positive and negative keywords respectively) in records where the patient wore either BTE or ITE aids are shown in Table 6. The association between these keywords and one or other type of hearing aid suggests the following: BTE aids were associated with high gain (amplification), e.g., be34, be36 and be52, and cases where changes had been made to the ear mould. ITE hearing aid types tended to use lacquer, had vents, required reshelling of ear impressions, had changes made to the hearing aid itself, were reviewed and the wearers were making progress. Some of the keywords in Table 6, were stemmed forms such as 'reshel' for 'reshell' and 'tinnitu' for 'tinnitus', since all the text was passed through Porter's stemmer [3] for the removal of grammatical endings. Some significant keywords were abbreviations, such as IMP for 'impression'. We did not use dictionaries to group terms with the same meaning (synonyms), instead using only the surface form of words because we wanted the procedure of keyword selection to be as automatic as possible. We assumed that the text notes used in the database were homogeneous, where the various technicians were consistent in their use of terminology and abbreviations. Although we did not do it, consistency in writing conventions could have been verified manually, since the identity of the technician is given in one of the record fields. Similarly, we assumed that two technicians treating the same patient would produce identical audiograms and choose the same hearing aid type. To do otherwise would have resulted in excessive subdivision of our data set.

**Table 6. Most significant positive and negative keywords in records with BTE/ITE aid [4]**

	Positive keywords	Negative keywords
BTE	mould, be34, map, gp, 92, audio, inf, be52, ref, staff, reqd, be36, contact	fta, reshel, appt, it, nn, nfa, 2001, rev, lacquer, hn, km, imp, review, 2000
ITE	fta, reshel, appt, it, nn, nfa, 2001, rev, lacquer, hn, km, imp, review, 2000, nh, vent, progress, aid, dt, taken	mould, be34, map, gp, 92, audio, inf, be52, ref, staff, reqd, be36, contact, tri, n, order

#### 4. LOGISTIC REGRESSION (LR) MODEL FOR ITE / BTE RIGHT EAR HEARING AIDS

Having determined that audiogram frequencies, gender, presence of tinnitus, use of tinnitus masker, age, mould and certain keywords were all associated with the decision between fitting a BTE or an ITE aid, we combined all these factors into a single logistic regression model. The input to the model was the matrix of patient data, where columns corresponded to attributes and rows corresponded to individual patient records. The output was a formula [5] in the form

$$L = \log(p / (1 - p)) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k. \quad (2)$$

In our case  $p$  is the probability that the patient should be fitted with an ITE aid, while  $(1 - p)$  is the probability that the patient should be given a BTE aid.  $b_0$  is a constant, and  $b_1$  to  $b_k$  are called the coefficients of the model. The values  $x_1$  to  $x_k$  are all either 1 or 0, depending on whether a given attribute in the patient's record is present or absent. The overall value  $L$  is greater than 0 if it is more likely that the patient should be given a BTE aid, while it is less than 0 if it is more likely that the patient should be given an ITE aid.

Before performing the logistic regression, we randomly divided the actual set of records into two parts, one containing 80 percent (5,736) of the records and the other containing the remaining 20 percent (1,433) of these records. We used the 80 percent subset as the training set for model construction, and the remaining records for testing the model. We performed this logistic regression on the records which had all fields filled for the right ear: AC (air conduction) and BC (bone conduction) thresholds, gender, age and text keywords (5,736 records), of which 128 also had non-null entries for diagnosis, 85 had non-null entries for masker, and 3982 had non-null entries for mould. This data was converted into discrete numeric values as inputs to the model, as follows: For AC thresholds below the first quartile (40db) a value of 0; for thresholds between the first and second quartile (55dB) a value of 1; for thresholds between the second and third quartile (75dB) a value of 2; and for thresholds above the third quartile a value of 3. The same method was used to assign values for BC thresholds and age, except in that the quartile thresholds were 25, 40 and 55dB and 60, 70 and 78 years respectively. The values for diagnoses were 0 for no tinnitus diagnosis and 1 for tinnitus. Finally, for gender, we assigned 0 for male and 1 for female. Regression coefficients and associated p values were found for all the model variables and those variables with p values more than 0.05 for the constant were discarded. Thus, we did not consider BC4000 (bone conduction at 4000Hz), age and diagnosis in the final model as the p values of their constants was more than 0.05, as shown in Table 7, Table 8 and Table 9.

Due to data sparseness, it was not possible to incorporate all the keywords discovered by an analysis of the free text into a single model. However, we did include a few significant keywords (meaningful words producing the highest chi-squared values, shown in Table 10), into the final model, along with all the categorical and numeric fields. These keywords were all acronyms where APPT stands for appointment, FTA for first time appointment, GP for general practitioner, MAP for processor

amplification map (associated with cochlear implants), NFA for no follow-up-appointment and REV for hearing aid review. In Table 10, the part of the model which takes into account the occurrence or otherwise of the selected keywords in deciding which type of hearing aid to suggest is given. Using keywords alone, the relative likelihoods of the patient needing an ITE or a BTE aid are given by equation (3)

$$\log[P(ITE / BTE)] = -0.16 + 0.06(APPT) - 0.77(FTA) + 0.62(GP) + 2.32(MAP) - 0.93(NFA) + 0.12(REV). \quad (3)$$

Table 7. Logistic regression for BC4000

	Regression coefficient b	Standard error se(b)	Z	P
Constant	-0.09	0.08	-1.12	0.26
Bc4000_ind1	-0.15	0.11	-1.33	0.18
BC4000_ind2	-0.20	0.09	-2.12	0.03
BC4000_ind3	0.09	0.09	1.01	0.31

\* Note: Bc4000\_ind1, Bc4000\_ind2 and Bc4000\_ind3 represent bone conduction threshold quartiles of 25, 40 and 55dB respectively.

Table 8. Logistic regression for age

	Regression coefficient b	Standard error se(b)	Z	P
Constant	-0.08	0.05	-1.49	0.13
Age_ind1	-0.13	0.08	-1.73	0.08
Age_ind2	-0.26	0.08	-3.48	0.00
Age_ind3	0.14	0.08	1.88	0.06

\* Note: Age\_ind1, Age\_ind2 and Age\_ind3 represent age quartiles of 60, 70 and 78 years respectively.

Table 9. Logistic regression for diagnosis

	Regression coefficient b	Standard error se(b)	Z	P
Constant	0.37	0.39	0.96	0.34
Diagnosis	-1.05	0.44	-2.37	0.02

Table 10. Logistic regression for keywords

	Regression coefficient b	Standard error se(b)	Z	P
Constant	-0.16	0.03	-5.63	0.00
APPT	0.06	0.15	0.37	0.71
FTA	-0.77	0.19	-4.05	0.00
GP	0.62	0.13	4.75	0.00
MAP	2.32	0.53	4.39	0.00
NFA	-0.93	0.32	-2.93	0.00
REV	0.12	0.10	1.12	0.26

As shown in Table 11 and Table 12 logistic regression coefficients for gender and masker were calculated. In Table 13, logistic regression values for air conduction (AC) at 250dB are given for each quartile and in Table 14, predicted log odds are

calculated using the regression coefficient values (b) from Table 13. Similarly, predicted log odds were calculated for AC500, AC1000, AC2000, AC4000, AC8000, BC250, BC500, BC1000, BC2000, gender, masker and mould.

**Table 11. Logistic regression for gender**

	Regression coefficient b	Standard error se(b)	Z	P
Constant	-0.23	0.04	-5.93	0
Gender	0.16	0.05	3.08	0

**Table 12. Logistic regression for masker**

	Regression coefficient b	Standard error se(b)	Z	P
Constant	-0.12	0.03	-4.65	0
Masker	-1.59	0.30	-5.25	0

**Table 13. Logistic regression for AC250**

	Regression coefficient b	Standard error se(b)	Z	P
Constant	-0.72	0.04	-17.23	0.00
AC250_ind1	0.54	0.07	8.15	0.35
AC250_ind2	1.29	0.07	17.26	0.00
AC250_ind3	2.18	0.12	17.91	0.00

\* Note: Ac250\_ind1, Ac250\_ind2 and Ac250\_ind3 represent Air conduction at 250dB quartile of 40, 55 and 75 dB respectively.

**Table 14. Predicted Log odds for AC250**

AC250 group	Logistic regression equation	Predicted log odds
0<AC250<=40	Log odds = $b_{\text{constant}}$	-0.72
40<AC250<=55	Log odds = $b_{\text{constant}} + b_{\text{AC250\_ind1}}$	-0.18
55<AC250<=75	Log odds = $b_{\text{constant}} + b_{\text{AC250\_ind2}}$	0.57
75<AC250	Log odds = $b_{\text{constant}} + b_{\text{AC250\_ind3}}$	1.45

To show how the model works, we take a sample record from the test set, as shown in Table 15. The attributes of this record are shown in the first column, and their values are shown in the second. Starting with a predicted log odds of 0 (meaning a BTE and ITE aid are assumed equally likely), we examine the values of the attributes in the record one by one, and add on the regression coefficient corresponding to that value of that attribute to our running total. The value of the first attribute, age is disregarded, but the next attribute gender has value “male”. For “male”, we calculated the regression coefficient of -0.23, so the running total becomes  $0 - 0.23 = -0.23$ . The relevant regression coefficients for each attribute are added in turn, ending with adding on -0.04 for the presence of the keyword “REV” in the free-text field. The final total of the regression coefficients is 9.98, which is our final log odds value, suggesting that it is much more likely that this patient would benefit most from a BTE hearing aid as opposed to an ITE aid.

**Table 15. Logistic regression – worked example**

Candidate variables (database record)	Actual values	Predicted log odds	Overall predicted log odds
Age	71	Not-significant	0
Gender	Male	-0.23	-0.23
AC250	75	0.57	0.34
AC500	70	0.72	1.06
AC1000	80	2.08	3.14
AC2000	90	1.19	4.33
AC4000	100	0.40	4.73
AC8000	100	0.09	4.82
BC250	40	-0.03	4.79
BC500	60	0.56	5.35
BC1000	65	0.56	5.91
BC2000	70	0.14	6.05
BC4000	70	Not-significant	6.05
Diagnosis	Tinnitus	Not-significant	6.05
Hearing aid type	BTE	To be found	6.05
Masker	No masker	-0.12	5.93
Mould	2107	4.09	10.02
Free-text words	REV	-0.16+0.12 = -0.04	9.98

Testing of our logistic regression models showed that overall there was 82.21% agreement between the predictions of our model and the actual hearing aid chosen by the audiologist (as given in the “type” field) as shown in Table 16. The agreement rate was higher for patients fitted with ITE aids (88%) than for those fitted with BTE aids (75%). The results were analyzed according to precision, recall and F-measures [6] using equations (5), (7) and (8) respectively, as shown in Table 18. For comparison, we performed a similar analysis using a Naïve Bayesian approach, and obtained 0.67, 0.76, and 0.71 as the precision, recall and F-score respectively for ITE and 0.66, 0.56 and 0.60 as the precision, recall and F-score respectively for BTE.

$$P = \frac{\text{Agreements\_of\_machine\_and\_human}}{\text{Total\_number\_in\_that\_category\_by\_machine}} \quad (4)$$

$$P_{ITE} = \frac{688}{(688 + 161)} = 0.81 \quad (5)$$

$$R = \frac{\text{Agreements\_of\_machine\_and\_human}}{\text{Total\_no\_in\_that\_category\_in\_reality}} \quad (6)$$

$$R_{ITE} = \frac{688}{(688 + 94)} = 0.88 \quad (7)$$

$$F_{ITE} = 2 \times \frac{P \times R}{P + R} = 2 \times \frac{0.81 \times 0.88}{0.81 + 0.88} = 0.84 \quad (8)$$

**Table 16. Overall results**

Results	Number of records	Percentage
Similar	1178	82.21
Not-similar	255	17.79
Total	1433	

**Table 17. ITE / BTE aid predicted results**

Machine results (Logistic regression model)	Human (actual data)		
	ITE	BTE	Total
ITE	688 (88%)	94 (12%)	782
BTE	161 (25%)	490 (75%)	651
Total	849	584	1433

**Table 18. ITE / BTE aid Precision, Recall, F-score**

	ITE	BTE
Precision	0.81	0.84
Recall	0.88	0.75
F-score	0.84	0.79

In Table 17, 782 and 651 are the counts of ITE and BTE aids respectively in the human-annotated test data, while 849 and 584 are the counts of ITE and BTE aids respectively in the machine predicted results. The overall agreement is much better than random (50%), but the performance of a classifier should also be compared against the “simplest possible algorithm” [7]. In our case, this would be to assume that all the patients should be assigned the more commonly prescribed type of hearing aid. In our test set 782 out of 1433 patients in the test set were given ITE aids, so simply assigning all the patients this type of aid would provide 54.6% agreement, which is referred to as the ZeroR baseline.

The theoretical upper bound of classifier performance is the inter-annotator agreement [5], in our case the rate at which two expert audiologists would assign the same hearing aid to the same patient. Unfortunately, we do not have data on this.

## 5. DISCUSSION & CONCLUSIONS

The associations between hearing aid type and audiogram type were confirmed by both the PCA/chi-squared and LR experiments described in this paper, and also by our previous work on associations between words found in the database and hearing aid type, and the previous findings by audiologists [1]. These approaches will form the basis for an audiology decision support system, where unseen patient records would be presented to the system, and the relative probability that the patient should be fitted with an ITE aid as opposed to a BTE hearing aid would be returned. The advantage of these techniques for the combination of evidence is that it is easy to see which variables contributed to the final decision.

Although our LR model did not find age as a significant factor, Meredith and Stephens [8] have found that the ITE hearing aid presents handling problems only in subjects over 75 years of age. Dillon [9] also found that BTE aids are easier to operate as they are larger in size, and thus would be more popular with older

people. The literature shows men and women preferring the two types for different reasons. Martin, et al. [10] found that more males choose ITE aids than females, because they perceive them to be a more advanced technology – though in reality the same makes and specifications are available in both styles, and neither model is more advanced than the other. They also found more females reporting that ITE aids are easier to handle than BTE. Mueller, et al. [11] found no difference in how embarrassed males and females feel about using a BTE aid. Our LR model did not include diagnosis (as mentioned above for Table 9), although we previously found [4] that there was a significant association between the choice of BTE hearing aid and not-tinnitus diagnosis (that is diagnosis other than tinnitus – ringing in the ear), by using chi-squared test. We also found, by using chi-squared test that BTE hearing aids were atypical of tinnitus-with-masker. Other factors mentioned in the literature which we could not test with our data are the greater cosmetic acceptability of the smaller ITE aids, comfort in wear, ease of use with spectacles, and sound quality [12].

We plan to validate our results by obtaining feedback from a professional audiologist, and by repeated sub-sampling of the data to produce the text set to find the unbiased estimation of model accuracy. A major advantage of both Naive Bayes and Logistic Regression is that they enable an explanation facility to be incorporated into any decision support tool, since it is easy to read back and see exactly which variables contributed exactly how much to the final decision of whether to fit a BTE aid or an ITE aid.

## 6. ACKNOWLEDGMENTS

We wish to thank Maurice Hawthorne, Graham Clarke and Martin Sandford at the Ear, Nose and Throat Clinic at James Cook University Hospital in Middlesbrough, UK, for making the large set of audiology records available to us.

## 7. REFERENCES

- [1] Stephens, S. D. G. 1984. Hearing-aid selection: An integrated approach. *British Journal of Audiology*. 18 (1984), 199-210.
- [2] Anwar, M. N., Oakes, M. P., Wermter, S., and Heinrich, S. Clustering Audiology Data. In *19th Annual Belgian-Dutch Conference on Machine Learning (BeneLearn 2010)*, Leuven, Belgium. (2010), [http://dtai.cs.kuleuven.be/events/Benelearn2010/submissions/benelearn2010\\_submission\\_7.pdf](http://dtai.cs.kuleuven.be/events/Benelearn2010/submissions/benelearn2010_submission_7.pdf) (Last accessed: 8<sup>th</sup> August 2011)
- [3] Porter, M. F. 1980. *An algorithm for suffix stripping*. Program, 14 (1980), No. 3, 130-137.
- [4] Anwar, M. N., Oakes, M. P., and McGarry, K. Chi-squared, Yule’s Q and Likelihood Ratios in Tabular Audiology Data. In *Electrical Engineering and Applied Computing*. Ao, S. L., and Gelman, L., Eds. Dordrecht: Springer Netherlands. 90, (2011), 465-376.
- [5] Altman, D. G. 1991. *Practical Statistics for Medical Research*. Chapman & Hall. 351-358, 403-404.
- [6] Manning, C.D., Raghavan, P. and Schütze, H. 2008. *Introduction to information retrieval*. Cambridge University Press. 142-144.

- [7] Manning, C. D., and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts, London, England. 233-234.
- [8] Meredith, R., and Stephens, D. 1993. IN-THE-EAR and Behind-The-Ear hearing aids in the elderly, *Scandinavian Audiology*. 22 (1993), 211-216.
- [9] Dillon, H. 2001. *Hearing aids*. Boomerang Press. 282, 284.
- [10] Martin, H., and Kane, S. 2008. Do NHS patients still want ITE aids? (Poster), South Tees Hospitals, NHS Trust. In *British Academy of Audiology Conference (UK)*. 5.
- [11] Mueller, G. H., and Budinger, A. C. 1990. Selection of hearing aid style. In *Hearing Instrumentation and Technology*. 2 (1990), 5-10.
- [12] Brooks, D. N. 1994. Some factors influencing the choice of type of hearing aid in the UK: Behind-The-Ear or In-The-Ear. In *British Journal of Audiology*. 28 (1994), 91-98.