INTERACTION IN SONIFICATION RESEARCH: A CITATION ANALYSIS OF THE 2004 ICAD PROCEEDINGS

Jason Holmes

School of Library and Information Science,
Kent State University,
Kent, Ohio
USA.
jholmes@slis.kent.edu

1. ABSTRACT

A citation analysis of ICAD 2004 proceedings was conducted to ascertain the leading (i.e., most cited) authors and information sources. The analysis found that S.A. Brewster is the most highly cited author and that previous ICAD proceedings were the most highly cited source. The observed citation behavior exhibited in ICAD 2004 compares nicely with the expected Lotkan and Zipfian distributions. Further citation analysis combining the previous ICAD conferences is necessary to get a fuller understanding of the ICAD corpus of literature. The issue of whether auditory display is emerging as its own discipline can be elucidated with a complete citation dataset. Continued efforts along these lines will contribute to our overall understanding of ICAD as a community of scholars and researchers

2. INTRODUCTION

Who are the highly cited authors in the ICAD community? What are the highly cited information sources in ICAD? What types of publications do ICAD researchers cite?

Bibliometrics is "the application of quantitative methods to the study of information resources."[1] As a subset of bibliometrics, citation analysis deals specifically with the frequency and pattern of citation in a given literature. Citation patterns can reveal influential authors and seminal works in a topic area. High citation frequency of either a particular author or a particular information source indicates that the author or information source is important to the field. Essentially, by describing and characterizing the citation practices of ICAD, we reveal something of the nature of the field itself.

Data are analyzed statistically to determine if they conform to recognized distributions for frequency of occurrence. The purpose of this analysis is to determine whether the set of documents cited are "normal" exhibitions of citing behavior. There are 2 mathematical distributions that suggest "normal" citing behavior; Lokta's distribution and Zipf's distribution.

2.1. Lotka's Distribution

In 1926 Alfred J. Lotka modeled the frequency distribution of scientific productivity and proposed a distribution that fit his observations.
"In the cases examined it is found that the number of persons making 2 contributions is about one-fourth of those making one; the number making 3 contributions is about one-ninth, etc.; the number making n contributions is about $1 / n$ squared of those making one; and the proportion, of all contributors, that make a single contribution, is about 60 percent."[2] To put it another way, a very small number of people will make a large number of contributions to the literature, while a large number of people will make only 1 contribution.

2.2. Zipf's Distribution

Originally working with occurrences of words in English, Zipf developed his distribution by counting the number of times a word occurred in a corpus of text and then rank ordering the list of terms. He observed that frequency of a term and that term's rank ordering have a surprisingly constrained relationship. "The name "Zipf's Law" has been given to the following approximation of the rank-frequency relationship: $rf = c$ where r is the rank of a word-type, f is the frequency of occurrence of the word-type, and c is a constant, dependent on the corpus." [3]

3. ASSUMPTIONS AND LIMITATIONS

Some basic assumptions must be made in order to proceed with this study. Primary among these is that citations reflect that use was made of previous research. It implies that at least part of a previous work has informed the current work.

Another assumption that must be made for the purposes of this study is that all citations are equal in weight. In other words, each citation from a source article will be treated as having made an equal contribution to the final product. This is clearly not the actuality of the production of a research paper, as some references will have greater influence on the finished work than others.

A third important assumption is that the authors of a work will cite all other works which contributed in some way to the research paper. Again, it seems self-evident that this is not entirely true. Some foundational or classic works may be so well known to the authors and to their audience that knowledge applied from those works is, at least in a given discipline, "common knowledge," and may not be cited. Additionally, material that the researchers read as background for a project and assimilated into their general understanding of a problem may not be included in references, since it is not seen as directly related to the paper as ultimately written.

The current study is limited in that it is only a snapshot of one year of ICAD conference activity. Data from additional previous conferences would no doubt give a better answer to the questions posed above.

4. DATA COLLECTION

Data for citation analysis were collected by recording the author, title, source, and year for each work cited in all of the papers in the ICAD 2004 proceedings as made available at the ICAD website [4] into a database for manipulation and analysis.

Data collection for citation analysis is neither simple not "clean." There is a good deal of variability that is difficult to fit into a standard format. Additionally, there is the practical matter that this type of data entry is expensive and absolute, ultra precision will not ultimately increase the value of the data set. Because were primarily looking at trends, a slight variation in data is acceptable. Consider these examples:

1. Proceedings of ICAD '97
2. Proceedings of the Fourth International Conference on Auditory Display
3. Proceedings of the 4th International Conference on Auditory Display

All three of these are textually distinct, but refer to the same proceedings of the same conference. Considerable effort is made to cleanup these sorts of inconsistency, but there will still be some (non-consequential) error.

5. ICAD 2004 PAPERS

The proceedings of the ICAD 2004 in Sydney, Australia contained 41 papers by 95 authors. The authors come from 11 different countries.

COUNTRY	FREQUENCY	% (n=95)
USA	35	37%
AUSTRALIA	22	23%
UK	11	12%
GERMANY	8	8%
CANADA	6	6%
FINLAND	5	5%
AUSTRIA	2	2%
ITALY	2	2%
MEXICO	2	2%
DENMARK	1	1%
FRANCE	1	1%

Table 1: Frequency of Author's Country

5.1. Authors cited

The proceedings of the ICAD 2004 in Sydney, Australia contained 41 papers. There were 1401 authors and co-authors cited in those 41 papers.

AUTHOR	FREQ.	% (n=1401)
BREWSTER, S.A.	25	1.8%
KRAMER, G.	22	1.6%
WALKER, B.N.	17	1.2%
HERMANN, T.	14	1.0%
BARRASS, S.	11	0.8%
SHINN-CUNNINGHAM	11	0.8%
VICKERS, P.	11	0.8%
ALTY, J.L.	9	0.6%
COOK, P.R.	9	0.6%
RITTER, H.	9	0.6%
EDWARDS, A.D.N.	8	0.6%
NEUHOFF, J.G.	8	0.6%
BRUNGART, D.S.	7	0.5%
GAVER, W.W.	7	0.5%
HUNT, A.	7	0.5%
PAI, D.K.	7	0.5%
VAN DEN DOEL, K.	7	0.5%
ARBOGAST, T.L.	6	0.4%
BAIER, G.	6	0.4%
HINTERBERGER, T.	6	0.4%
WENZEL, E.M.	6	0.4%

Table 2: Frequently Cited Authors

Lotkan Distribution			
Contributions	Authors	Predicted	Differences
25	1	1.15	0.15
22	1	1.49	0.49
17	1	2.49	1.49
14	1	3.67	2.67
11	3	5.94	2.94
9	3	8.88	5.88
8	2	11.23	9.23
7	5	14.67	9.67
6	4	19.97	15.97
5	12	28.76	16.76
4	22	44.94	22.94
3	39	79.89	40.89
2	102	179.75	77.75
1	719	549	170

Table 3: Lotka Distribution of Author Citations
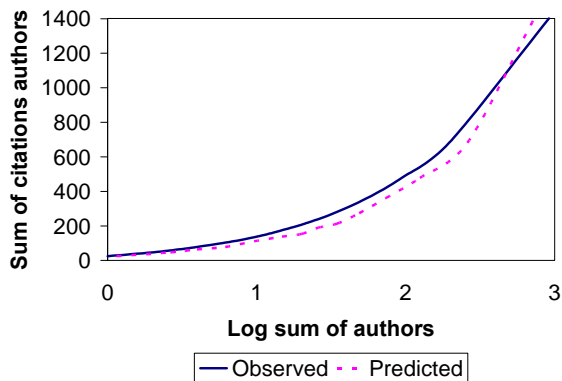
Figure 1: Lotka Distribution of Author Citations

Zipf Ranks			
Cohort	Rank	Frequency	Product
Size	r	f	rf
1	1	25	25
1	2	22	44
1	3	17	51
1	4	14	56
3	6	11	66
3	9	9	81
2	11.5	8	92
5	15	7	105
4	19.5	6	117
12	27.5	5	137.5
22	44.5	4	178
39	75	3	225
102	145.5	2	291
719	556	1	556

Table 4: Zipf Ranks of Author Citations

5.2. Publications cited

TITLE	FREQUENCY	% n=754
PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON AUDITORY DISPLAY	107	14.2%
JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA	51	6.8%
AUDITORY DISPLAY: SONIFICATION, AUDIFICATION, AND AUDITORY INTER...	20	2.7%

INTERACTING WITH COMPUTERS	13	1.7%
JOURNAL OF THE AUDIO ENGINEERING SOCIETY	8	1.1%
JOURNAL OF EXPERIMENTAL PSYCHOLOGY: ALL	8	1.1%
NATURE	7	0.9%

Table 5: Frequently Cited Sources

Lotkan Distribution			
Contributions	Sources	Predicted	Differences
107	1	0.03	0.97
51	1	0.14	0.86
20	1	0.94	0.06
13	1	2.23	1.23
8	1	5.89	4.89
7	1	7.69	6.69
6	2	10.47	8.47
5	2	15.08	13.08
4	7	23.56	16.56
3	10	41.89	31.89
2	42	94.25	52.25
1	377	267.6	109.4

Table 6: Lotka Distribution of Source Citations
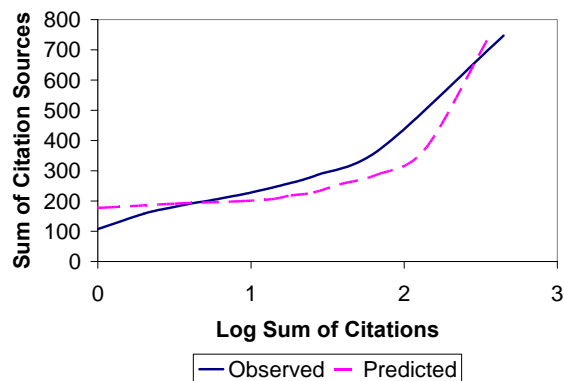
Figure 2: Lotka Distribution of Publications Cited

Zipf Ranks			
Cohort Size	Rank	Frequency	Product
	r	f	rf
1	1	107	107
1	2	51	102
1	3	20	60
1	4	13	52
1	5	8	40
1	6	7	42
2	7.5	6	45
2	9.5	5	47.5
7	14	4	56
10	22.5	3	67.5
42	48.5	2	97
377	258	1	258

Table 7: Zipf Ranks of Source Citations

5.3. Publication Types

Type	Frequency	% (n≈781)
PERIODICALS and PROCEEDINGS	472	60%
BOOK	169	22%
WEBSITE	70	9%
DISSERTATION	23	3%
OTHER/NA	47	6%

Table 8: Publication Types

5.4. Years cited

Year	Count	% (n=735)
2002	94	13%
2003	93	13%
2001	61	8%
2000	53	7%
1994	53	7%
2004	43	6%
1997	40	5%
1999	36	5%
1998	31	4%
1996	31	4%
1995	28	4%
1993	22	3%
1992	16	2%
1990	15	2%

Table 9: Years of Publications Cited

6. STATISTICAL ANALYSIS

Regression analysis on the Zipf ranks of cited authors (table 4) returns an R-squared value of 1. This would indicate that the distribution of authors cited in ICAD 2004 fits very well with the predicted Zipfian model. Regression analysis on the Zipf ranks of source citation (table 7) returns an R-squared value of .93. This indicates that although the source citation does not have exactly Zipfian characteristics, it is pretty close to the expected distribution.

7. DISCUSSION AND CONCLUSION

The International Conference on Auditory Display is indeed international as indicated in table 1.

Citation analysis reveals the most cited author in the ICAD 2004 proceedings to be S. A. Brewster. Not surprisingly, the most cited source (by far) is previous ICAD conference proceedings. This would suggest that auditory display, as a discipline, is emerging. By relying less on "outside" or "interdisciplinary" sources, ICAD is drawing from its own literature.

Further citation analysis combining the previous ICAD conferences is necessary to get a full understanding of the ICAD corpus of literature. The issue of whether auditory display is emerging as its own discipline can be elucidated with a complete citation data set. Continued efforts along these lines will contribute to our overall understanding of ICAD as a community of scholars and researchers.

8. REFERENCES

[1] D.P. Wallace, "Bibliometrics and Citation Analysis." In *Principles and Applications of Information Science for Library Professionals*. J.Olsgaard, ed. ALA,1989, pp.10–26.

[2] A. J. Lotka, "The Frequency Distribution of Scientific Productivity," *Journal of the Washington Academy of Sciences*, vol. 16, no. 12, June 19, 1926, pp. 317-323.

[3] Ronald E. Wyllys, "Empirical and Theoretical Bases of Zipf's Law," *Library Trends*, vol. 30, no.1, Summer 1981, pp. 53-64.

[4] ICAD 2004: The 10th Meeting of the International Conference on Auditory Display. Proceedings.[Online]. Available: http://www.icad.org/websiteV2.0/Conferences/ICAD 2004/program.htm