

順位投票에 의한 위원회 결정의 公正性 검정*

— 동아 음악 콩쿠르에 의한 분석 —

林正德** · 蔣永在***

< 目 次 >

- I. 서론
- II. 주관적 채점에 의한 순위결정의 문제점
- III. 실증분석
- IV. 요약과 결론

I. 서 론

경쟁은 優劣을 가리는 과정이다. 대부분의 경기는 일정한 규칙 하에서 힘이나 기술을 써서 누가 더 잘하는가를 가릴 수 있다. 시장에서는 경제원리가 작용하여 優劣을 판가름한다. 그러나 어떤 경기나 시험 또는 경쟁에서는 심판 또는 심사위원이 기량이나 기술을 주관적으로 판단하여 等位를 정하게 된다. 몇 가지 예를 들면 운동경기 중 다이빙이나 체조, 한국의 사법 혹은 행정고시, 각종 기능대회 등은 그런 방식에 의해 경쟁한다. 이런 것들 중에서 예술 분야의 경쟁이 가장 많이 또 광범위하게 심사위원의 판단에 의한 경쟁에 의해 순위를

* 이 논문은 1995년 2월 한국경제학회에 처음 발표되었던 논문이다. 저자들은 학회의 논평자(주만수 교수)와 2명의 심사위원에게 감사를 드린다. 논평과 지적사항들이 본 논문을 수정, 보완하는 데 큰 도움이 되었다.

**釜山大 貿易學科

***仁濟大 經濟學科

결정하는 방식을 택하고 있는데 아직까지는 그보다 더 나은 다른 代案이 발견되지 않고 있다.

심사위원의 주관적 판단에 따른 順位決定方式은 몇 가지 중대한 문제를 야기한다. 먼저 심사위원들간의 의견이 전체적으로 일치하였는가의 문제이다. 개별적 행동여부에 관계없이 심사위원의 의견이 일치하지 않았다는 통계적으로 유의한 결과가 나타나면 신뢰성에 문제가 제기될 수 있다. 다음으로, 누가 심사위원이 되는가에 따라 결정내용이 달라질 수 있다. 이것은 따로 설명을 필요로 하지 않는 근본적인 문제인데 우리 사회에서 의외로 크게 問題視되지 않고 있다. 셋째, 심사위원의 훈련과 성장배경 및 개인적 취향에 따라 전혀 다른 판정을 할 가능성이다. 예를 들어 음악이나 미술 경연에 있어서 어떤 계파나 흐름에 속하거나 어떻게 훈련받았는가에 따라 곡이나 작품에 대한 해석이나 평가가 달라질 수 있고, 개인적인 예술관이나 인생관에 의해 평가의 격차는 더 벌어질 수도 있다. 사법시험의 경우에도 사정은 마찬가지이다. 설사 학계에서 공통적으로 인정되는 학설이나 해석이 있다 하더라도 채점자에 따라서 강조하는 점이 다르거나 선호점이 다르면 미세한 격차가 순위결정이나 당락에 결정적인 영향을 미칠 수 있다. 이 주장은 대학입시를 포함한 모든 주관식 채점방식에 의한 경쟁에 대한 일반적인 문제로 확장될 수 있다.

넷째, 앞의 두 문제가 해결된다 하더라도 심사위원간의 명시적 또는 묵시적 담합이나 의도된 또는 무의도적 조작가능성이다. 예를 들면 심사위원이 개인 혹은 그룹으로서 어떤 의도된 판정을 시도한다면 순위결정에서 변동이 얼마든지 일어날 수 있는데 음악콩쿠르나 미술전 같은 경우에 1位와 2位는 천양지차의 결과를 가질 수도 있다. 이런 담합이나 조작은 사전에 전혀 시도되지 않았다 하더라도 같은 흐름의 훈련을 받았다거나 평소 취향이나 테크닉이 비슷할 경우에는 무의도적인 담합이 이루어질 수도 있다.

판정의 공정성 문제는 끊임없이 제기된다. 그런데 여러 가지 문제에도 불구하고 사회의 여러 분야에서 순위결정(채점)방식의 경쟁은 계속되고 있다. 이런 문제가 너무 심하다면 순위결정방식에 의한 경쟁제도를 폐지하면 된다. 그러나 현실적 측면에서 수요와 공급측 모두가 이 제도에 의한 경쟁을 원하고 있으므로 이런 종류의 순위결정은 더 늘어나지 결코 줄어들지는 않을 것이다. 따라서 현행제도를 중심으로 이 방식의 공정성을 분석하고 검토해 보는 것이 필요할 것이다.

본 연구는 동아음악콩쿠르의 채점결과를 Kendall 일치계수검정과 Kendall 순위상관계수방법으로 분석하여 공정성을 검증해 보고자 한 것이다. 동아음악콩쿠르는 오래 전부터 심사위원의 채점표를 공개하므로 자료 입수가 가능하다. 이 콩쿠르 중 참가자와 심사위원이 가장 많은 분야가 피아노 부문이므로 이 부문만을 본 연구대상으로 하였다. 연구는 심사위원의 선정숫자에 의한 영향력변화, 위촉빈도와 위촉대상에 따르는 공정성의 영향가능성, 심사의 전반적 신뢰성 측정, 개인 및 그룹으로서의 의도적 또는 무의도적 담합가능성 등을 통계적 방법으로 검증해 보려고 노력하였다.

모든 자료는 동아일보가 지상에 발표한 심사결과를 이용하였으며 심사위원 개인의 노출을 피하기 위해 심사위원 이름은 열코드와 번호에 의해 처리하였으므로 논문에 일체 나타나지 않는다.

자료만 제공된다면 사법, 행정고시를 비롯한 각종 고시나 예술분야의 경연 또는 모든 경쟁을 통한 시험방법에 이러한 방법을 이용하여 공정성이나 효율성 검증을 할 수 있을 것이다.

II. 주관적 채점에 의한 순위결정의 문제점

대부분의 교육이나 경쟁은 평가과정을 거치게 된다. 대학의 학점이나 중고등학교의 성적과 같이 평가를 그룹으로 하여 개별부문(과목)의 순위를 결정하지 않는 방식도 있으나 예술이나 문학, 콩쿠르, 고시 등과 같이 순위를 결정하는 방법에 의해 합격이나 결과를 평가하는 경우도 허다하다. 학교의 성적이 진학이나 취업에 중요한 영향을 미치는 것과 마찬가지로 콩쿠르나 고시의 결과는 더 직접적이고 결정적인 영향을 참가자에게 미친다. 고시의 합격여부는 설명의 필요조차 없고 콩쿠르의 순위결정도 참가자의 진로와 장래에 큰 변수로 작용한다. 예를 들면 동아 음악콩쿠르의 대상 수상자는 개인적인 영예와 입상경력 외에도 병역이 면제되고 해외연수 특전을 부여하는 등의 엄청난 혜택이 부여되었으므로 경쟁순위는 매우 중요한 역할을 한다. 고시의 경우는 일정 정원까지 합격자를 결정하므로 순위가 음악콩쿠르만큼 중요하지 않을 수가 있으나 콩쿠르의 경우에는 1위와 2위와의 차이는 천양지차가 있어서 혜택이나 인정의 정도에 결정적인 영향을 미치게 된다.

교육학에서 평가의 문제가 중요하기 때문에 많은 연구가 있으나 위와 같은

관점의 문제를 다루고 있는 경우가 거의 없고 통계학에서도 이런 부문에 관한 연구는 없는 것으로 조사되었다.¹⁾ 그러나 연구가 없었다고 문제가 없었다는 의미는 아니다. 본 연구에서 대상으로 하는 음악콩쿠르의 경우 참가자들이나 주최자들이 運이나 기회로 돌려버리고 마는 부분에 의도적이거나 비의도적인 결합이 있었을 가능성이 있으므로 이런 분석을 통해 그 가능성을 검증할 필요가 있는 것이다. 이 문제는 질적인 것에 대한 평가를 수반하므로 어려운 점이 있으나 통계적 처리를 통해 어느 정도까지 분석이 가능하다. 이 방법은 측정문제로 난관에 봉착해 있는 사회후생의 간접적인 측정에도 시사점이 될 것이다.

피아노 연주를 한번만 듣고 연주자의 자질이나 기술, 음악성 등을 대체적으로 판정할 수 없는 것은 아니겠지만 콩쿠르에서 예선을 거쳐 본선에 진출한 재능있는 연주자들의 연주에 대해 순위를 정한다는 것은 어려운 작업이다. 특히 곡의 해석, 연주의 기법이나 방법이 하나로 통일되어 있거나 음악가(피아니스트)를 양성하는 기관이나 조직이 하나뿐이라 하더라도 개인의 개성이 크게 작용할 수 있는데 현재와 같이 이런 요인들이 나라별로, 지역별로, 학교별로, 학과별로 나누어져 있는 경우에는 객관적인 평가가 불가능할 수도 있다. 거기에다 개인적인 개성이나 취향이 추가되므로 심사위원의 심사당일 신체 및 기분 조건 또 개인적인 조작의 가능성까지 포함되면 문제는 상당히 복잡해진다.

그렇다고 음악계에 어떤 큰 흐름이 없는 것은 아니다. 경제학과 마찬가지로 음악계에도 주류와 비주류의 큰 흐름이 있을 수 있고 주류는 경제적으로 가장 부강한 나라에서 일반적으로 형성된다. 서양음악이 유럽에서 꽃피고 발전했지만 현재 연주나 교육의 중심지는 미국인 것과 마찬가지로 대체적인 흐름을 결정하는 기준이 있을 수 있다. 그러나 세계적인 음악콩쿠르가 모두 미국에 집중되어 있는 것은 아니다. 분명한 것은 어떤 탁월한 연주자가 어느 콩쿠르에 가서도 같은 성적을 낼 수는 없다는 사실이다. 그것은 각 콩쿠르가 가지고 있는 특색이나 취향이 있을 수 있고 오랜 기간동안에 그것이 전통화되어서 그 규격이나 취향에 맞는 연주가 더 인정을 받게 된다는 의미이다.

한국의 경우에는 서양음악이 과거부터 수입되었고 본격적으로 발전한 것은 경제발전과 궤를 같이한다고 볼 수 있다. 지금 서양음악계를 주도하는 대부분

1) 교육학에서의 평가에 관련된 논문과 Biometrika를 중심으로 한 문헌조사에서는 이와 유사한 연구가 발견되지 않았다.

의 지도자나 연주자들은 미국이나 유럽에 가서 훈련을 받고 온 사람들이다. 시대를 초월하는 재능이나 기술이 가끔씩 출현하지 않는 것은 아니지만 연주의 질적인 평가는 심사위원의 훈련과정 및 성장배경과 직간접의 관계를 맺을 수밖에 없다. 미국과 유럽 및 한국의 차이, 또 한 국가 내에서의 지역, 학과, 학교의 차이와 심사위원의 개성이 심사결과에 영향을 미치게 되어 있다. 그러나 해당 심사위원은 아마도 이런 지적을 부인할 것이고 이것은 음악의 질적인 측면과 관계되므로 어떤 것으로도 반박하거나 결함을 지적하기는 어렵다.

이 연구에서 밝히고자 하는 부분은 질적인 측면과는 관계없는 어느 정도 객관적인 측정이 가능한 부분이다. 즉 연주의 질적 평가는 분석할 수 없지만 공정성과 관계 있는 본질적인 문제를 다루고자 한다. 먼저 어떤 연주자의 연주에 대해 모든 심사위원이 똑 같은 판정을 했다면 담합행위가 없었던 이상 그 평가는 공정했다고 할 수 있다. 물론 왜곡되었으나 일관성 있는(biased but consistent) 결과일 수도 있다. 어떤 학과나 계열적인 일관성일 수도 있다. 즉 이때에는 음악성의 一方性이 나타난다. 그러나 심사의 측면에서는 객관적인 타당성이 있었다고 볼 수 있다.

두번째의 극단적인 경우는 심사위원의 판정결과가 심사위원마다 다른 결과이다. 이때에는 심사위원간의 독립성은 증명되는 측면이 있지만 심사의 객관적 공정성은 전혀 보장되지 않는다. 이 경우 콩쿠르의 존재 의미가 퇴색되며 순위는 우연적인 요인에 의해서 결정된다. 이런 가능성이 전혀 없는 것은 아니고 현실적으로도 가끔씩 나타난다.

세번째는 앞서의 두 가지의 중간에 위치하는 형태이다. 어떤 심사위원들은 같거나 비슷한 성향의 결과를 나타내나 다른 심사위원은 개별적으로 혹은 집단적으로 (전혀) 다른 평가를 내리는 경우이다. 긴 설명이 필요없이 어느 한 해의 심사결과를 가지고 예를 들어보자. <표 1>은 1973년도 동아음악콩쿠르 피아노부의 채점표이다.

<표 1>

심사위원 연주자	A	B	B	C	D	E	F	G	H	I	J	K	L	M	N	합 계	등 위
가	3	1	1	2	2	4	1	2	4	3	1	3	2	2	·	31	2
나	1	3	2	1	4	3	2	1	1	1	4	2	1	1	·	27	1
다	2	2	4	4	1	1	4	4	2	2	3	1	3	3	·	36	3
라	4	4	3	3	3	2	3	3	3	4	2	4	4	4	·	46	4

표에서 보는 바와 같이 그 해의 1位 입상자는 출석 심사위원 14명 중에서 7명으로부터 1位점수를 받았으나 3명에게는 2位, 2명에게는 3位, 또 다른 2명으로부터는 4位(최하위)의 판정을 받았다. 이 경우 심사위원이 조금 다르게 구성되면 순위가 다르게 결정될 가능성이 충분히 있다. 또 같은 심사위원이라도 연주자간의 순위를 조금만 다르게 하면 전체 순위가 달라질 수 있다.

이하에서 실증적으로 검증하려는 부분은 이런 문제들이다. 콩쿠르라는 공개 경쟁이 있는 한 판정의 객관성은 보장되어야 하므로 통계적 유의성 검증방법을 이용하여 테스트할 수 밖에 없다. 검정해 보고자 하는 가설은 다음과 같다.

1. 심사위원간의 의견일치 정도

심사위원이 개인 혹은 그룹으로서의 독립적 행동을 하여 심사결과에 영향을 미쳤는가를 알아보기 전에 심사위원간의 의견이 전체적으로 일치하였는지 아니면 유의적으로 일치하지 않았는지를 검정해 보아야 한다. 이때 개별심사위원의 행동은 별도로 분리해 내지 않으면서 전체적인 일치성을 테스트할 수 있다.

2. 심사위원의 개인적 역할

가) 심사위원들이 개인적으로 독립적으로 행동한 결과 전혀 다른 순위의 판정을 하였다면 그 콩쿠르는 우연에 의해서 결정되고 누가 심사위원이 되느냐에 따라서 결과가 계속적으로 달라질 수 있다. 이 경우는 콩쿠르의 存在目的에 의문을 제기할 수 있다. 이 때에는 누가 심사위원에 자주 위촉되느냐가 판정에 제일 중요한 영향을 미치게 될 것이다. 연관된 의문은 심사위원의 교체정도와 판정결과에 유의적 차이가 있는가이며 또 동아콩쿠르의 경우 시기적으로 심사위원교체에 특색이나 區分이 있었는가에 대한 검증의 필요가 있다.

나) 일부 또는 전 심사위원들이 판정에 어떤 성향을 보이는 경우이다. 이 경우에는 판정결과와 객관성이라는 면은 부각된다. 이 경우에도 개별 심사위원의 행동이 전체 판정에 어느 정도 영향을 끼치는가는 중요한 관심사이다. 만약 개별 심사위원이 최종 심사결과와 같거나 비슷한 채점을 하였다면 공정성은

일단 보장된 것으로 볼 수 있다. 만약 개별 심사위원이 최종 결과와 다른 채점을 하였고 이 위원이 심사결과에 결정적인 역할(pivotal role)을 하였다면 공정성의 차원에서는 문제가 제기된다.

부연해서 말할 점은 공정성과 음악성은 별개의 문제라는 점이다. 어떤 심사위원이 최종결과와 정반대의 심사를 했다고 해서 그 사람의 음악성이 다른 사람에 비해 떨어진다고나 자질이 열등하다는 등의 판단은 할 수 없다. 오히려 그 위원만이 올바른 음악관이나 판단력을 가졌다고 주장할 수도 있다. 문제는 콩쿠르라는 공개경쟁의 과정에서 소위 말하는 객관적 공정성이 결여되는 경우의 우연성이며 전체 성향과 다른 위원이 심사결과에 어느 정도 영향을 미치는가 하는 점이다.

3. 그룹으로서의 역할

가) 심사위원이 통계적으로 독립적인 행동을 하지 않는다면 자신의 실제 의사와는 관계없이 그룹에 의한 행동을 하게 된다. 이 경우 그룹은 최대 $n-1$ 개 존재할 수 있으나 대체적으로 최종결과에 제일 큰 영향을 미치는 그룹(이하 主流)과 최종결과와 상반되는 판정을 한 그룹(이하 非主流)으로 大別된다. 재삼 지적하고자 하는 점은 최종 결과에 결정적 영향을 미치는 그룹이 음악적으로 가장 우수한 집단이라고는 할 수 없다. 그러나 경쟁의 성격상 최종결과와 다른 행동을 하는 위원이 개별적으로 혹은 집단적으로 어떤 성향을 나타내는가라는 물음은 필요하다. 이 물음은 최종결과와 비슷한 행동을 하는 위원 그룹에게도 같이 적용된다.

나) 개별위원의 경우와 마찬가지로 非主流의 행동이 심사결과에 미치는 영향 가운데서 가져볼 수 있는 가정은 만약 非主流 가운데서 일부가 심사위원으로 위촉되지 않았거나 실제 심사에 참가하지 않았더라면 결과에 어떤 차이가 나타났을까 하는 것이다. 다른 면으로는 主流 非主流에 관계없이 비슷한 심사 경향을 가지는 위원들이 그룹으로서의 영향력을 행사하는 경우가 있는가 하는 점이다.

다) 또 다른 측면은 심사위원으로 자주 위촉되는 위원들이 심사 때마다 어떤 일정한 성향을 나타내는가 또는 그룹적인 행동을 하는가 하는 점이다. 이런 현

상은 앞서 지적한 훈련배경, 음악성 등과 상관관계를 가질 수도 있고 다른 요인이나 혹은 우연적인 요인에 영향을 받을 수도 있다. 그러나 이런 성향이 존재한다는 것은 판정과 직접적인 관계가 있으므로 究明해 보아야 할 점이다.

4. 담합의 가능성

2人 이상의 위원이 있을 경우 담합의 가능성은 항상 존재한다. 담합의 방법은 명시적이거나 묵시적인 의사전달이나 계약에 의한 직접적인 것일 수도 있고 출신배경이나 음악성의 공통성 혹은 연주상의 테크닉 및 곡해석의 일치 등 여러 가지 형태로 존재할 수 있다. 또 이런 경우의 담합은 본인들이 전혀 의식하지 못한 채 결과에 의해 나타날 수도 있다. 통계적 방법은 담합의 사실성 여부를 가려낼 수 없고 실제 결과에 의해 그 가능성을 제시할 수 있다. 보다 노골적인 형태로 연주자와의 직접 담합도 그 가능성은 있다. 연주자가 심사위원으로 위촉될 가능성이 있는 전문가를 찾아가서 특별레슨이나 지도를 받는 경우에는 연주자와의 담합이 가능하다. 그러나 이 경우는 위원간의 담합이 아니므로 개인적 역할에서 나타날 수 있는 가능성이다.

Ⅲ. 실증분석

1. 심사위원의 구성과 교체율

심사위원들이 담합적인 행동을 하지 않는다고 하면 심사위원의 수가 많을수록 또 유자격 심사위원의 교체율이 높을수록 심사의 공정성은 상승한다고 할 수 있다. <표 1-1>의 첫 열은 1966년부터 1991년까지의 심사위원으로 위촉된 상황이다. 심사위원의 수와 교체율에 따라 5기간으로 나누어 볼 수 있는데, 즉 1966년부터 1969년까지, 1970년부터 1975년까지, 1976년부터 1980년까지, 1981년부터 1990년까지 그리고 1991년이후이다. 1966년부터 1969년까지는 심사위원의 수가 고정되어 있고 교체율도 매우 낮다.

1970년에서 1975년의 기간에는 위원수가 약 36% 이상 증가하고 심사위원도 그 전 기간에 비해 상대적으로 더 많이 교체되었다. 1976년부터 1980년까지는

대체적으로 심사위원의 수가 증가하기 시작하여 1981년부터 고정되기 시작하였다. 1974년에 심사위원의 교체율이 40%에 달해 상당한 물갈이가 있었고 이후 그 전보다는 높은 교체율이 계속되었다. 1981년부터 1990년까지는 심사위원 수가 증가된 채로 안정적이며 교체율은 큰 변동이 없다가 1990년도에는 교체율이 50%에 달하였다. 1991년에는 심사위원의 수가 감소하고 교체율이 94%이상에 달해 새로운 변화가 시도되었다고 볼 수 있다.

이상의 시간적 추이를 감안할 때 주최측이 나름대로 심사의 공정성이나 음악성을 높이기 위해 노력을 해오고 있었다고 해석할 수 있다. 이런 의도가 실제의 결과에서 공정성 제고나 우연성의 감소로 나타났는지는 별도로 검증해보아야 할 과제이다. 심사위원의 수가 무작위적으로 증가하면 담합이 없는 경우 대체로 공정성이 증가한다고 볼 수 있다. 또 무작위적 교체율이 높아도 공정성이 올라갈 수 있다. 그러나 주최측이 심사위원 선정기준을 발표한 적은 없고 교체가 무작위적으로 이루어진 증거는 위에서 발견할 수 없다.

〈표 1-1〉 심사위원 구성과 교체율

연도	본선진출자	위촉심사 위원수	참가 심사 위원수	심사위원수 (증감률)	새심사위원 (구성비)
1966	6	11	11		
1967	2	11	11	0.000	0.182
1968	5	11	11	0.000	0.000
1969	3	11	11	0.000	0.091
1970	4	15	14	0.364	0.267
1971	3	15	15	0.000	0.067
1972	2	15	15	0.000	0.067
1973	4	15	14	0.000	0.000
1974	2	15	14	0.000	0.400
1975	3	15	14	0.000	0.067
1976	3	16	15	0.067	0.188
1977	3	16	12	0.000	0.188
1978	3	17	15	0.063	0.294
1979	3	19	16	0.118	0.105
1980	5	20	18	0.053	0.150
1981	2	23	22	0.150	0.130
1982	4	23	19	0.000	0.174
1983	3	23	17	0.000	0.174
1984	4	23	20	0.000	0.174
1985	2	24	21	0.043	0.208
1986	4	24	21	0.000	0.042
1987	5	24	20	0.000	0.292
1988	6	23	20	0.000	0.435
1989	7	22	18	-0.042	0.138
1990	5	24	20	-0.043	0.500
1991	5	18	15	-0.250	0.944

기간별로 볼 때 시간의 경과에 따라 심사위원으로 위촉되는 수가 증가했다는 면에서는 바람직하다고 하겠다. 즉 평균의 법칙에 의해 표본의 수가 증가하면 평균의 분산은 감소한다. 심사의 공정성이라는 면에서 보면 새로운 심사위원의 구성이 높을 때가 바람직하다. 예를 들면 이들 새로운 심사위원들에게 심사에서의 특별한 배려를 바라고 개인교습을 받으려고 참가자가 접촉을 시도하기는 어려울 것으로 추정할 수 있다. 새 심사위원의 구성비가 25%를 넘는 연도를 보면 1970년, 1974년, 1987년, 1988년, 1990년, 1991년 등이다. 또한 이들 연도에서도 그 구성비가 대체로 증가하고 있음을 알 수 있다. 1991년은 새 심사위원 구성비가 94%를 넘고 있어 특기할만 하다.

〈표 1-2〉는 1966년부터 1991년까지 각 심사위원이 심사위원으로 위촉된 빈도수를 나타낸다. 앞에서 설명한 대로 각 심사위원은 코드번호로 표시되어 있다.

〈표 1-2〉 심사위원으로 위촉된 회수²⁾

빈 도	심사위원	빈 도	심사위원
23	E2(20) A6	9	F6(5) G5
21	I6(15)	8	F1 C4 K5(7) I2 H6
19	C1 D6	7	F3(6) G1 J1 D1(5)
17	D5	6	E1 C5 B1 J4
16	D3 F4 B6 J5(12)	5	G2 E6(4) E4
13	F5(12)	4	I4 J4 E3(3) B4 K1 J3 A3
12	I5	3	G3 E5 K3(2) I1 F2(2) J2(2)
11	G4(8) H4(10)	2	H2 B2 K6 D2 A5 H3 A1
10	H1(4) G6 H5(9)	1	B5 C6 B3 D4 J6 C3 C2
			A4 I3 A2 K2

총 26회 중 연 66명의 심사위원이 위촉되었는데 그 중 2명이 최고 23회까지 위촉되었다. 즉 이 두 위원은 1991년의 전면 물갈이 이전까지 거의 고정적으로 위촉되었다. 10회 이상 위촉된 위원도 17명에 달하고 있고 1991년을 제외하면 대부분의 위원이 최소 2회 이상 위촉되고 있다. 이런 현상은 콩쿠르의 특성유지, 심사위원 자원수의 제한, 특정 목적이나 기준에 의한 선발 등 여러 가지로 해석될 수 있는데 선발성향으로 보아 어떤 의도 하에서 위촉된다고 보아야 하며 주최측이 무작위적으로 선정하지는 않는다고 할 수 있다. 질적인 측면이

2) ()안의 수는 실제 심사에 참여한 회수를 나타낸다.

在되므로 이런 방법의 善惡을 논하는 것은 문제가 된다. 그러나 이것이 公正性과 직, 간접적으로 연결되는가는 자료에 의해 검증해 보아야 한다.

2. Kendall 일치계수검정(Kendall's Concordance Test)

먼저 해 볼 수 있는 公正性 검정방법으로 콩쿠르에 참가한 연주자들에 대해 등위를 부여한 심사위원들의 심사결과를 이용하여 심사위원들간에 의견이 일치하는가를 검정한다. Kendall의 일치계수검정(Kendall's Concordance Test)에 의해 다음의 귀무가설과 대립가설을 설정할 수 있다.

H_0 : k명의 심사위원들간에 의견이 일치하지 않는다.

H_1 : k명의 심사위원들간에 의견이 일치한다.

일치계수검정량 W 는 0과 1사이의 값을 가지며, $W=0$ 이면 심사위원들간에 의견이 완전히 불일치하는 경우이며 $W=1$ 이면 심사위원들간에 의견이 완전히 일치하는 경우이다.

콩쿠르 참가자수가 3명 이상 7명 이하이고 심사위원의 수가 3명 이상 20명 이하인 경우에는 소표본의 유의수준에 해당되는 임계치를 사용해야 한다. 반면 참가자수와 심사위원수가 더 많은 경우에는 χ^2 검정량으로 변형된 통계량을 이용하여 검정한다.

다음의 <표 2-1>은 켄달일치계수를 피아노 콩쿠르에 적용해 계산한 결과이다. 21년간의 결과를 보면 5번은 구성된 심사위원들간에 의견이 일치한다고 할 수 없는 경우로 나타났다.

1969년, 1971년, 1975년, 1977년, 1979년의 경우에 심사위원들간의 의견이 통계적으로 일치하고 있지 않다. 21회 중 5회의 빈도로 심사위원들간의 의견불일치가 발생하고 있다는 사실은 심사위원의 구성에 문제가 있었다는 증거가 될 수 있다. 그런데 이러한 의견의 불일치는 1980년 이후에는 나타나고 있지 않다는 것이다. 이 결과는 두 가지 의미를 동시에 내포한다. 한편으로 보면 심사위원들이 의견일치가 되도록 구성되어 가고 있는 바람직한 면이 있다고 하겠다. 더불어 제기되는 문제는 의견이 다른 심사위원들이 1980년 이후로 심사위원에서 제외되거나 심사위원들이 시간이 흐름에 따라 동질화되어가거나 담합하는 경향을 나타낼 가능성을 보여준다. 심사위원의 교체율을 <표1-1>에서 검토해 보면 심사위원의 교체율이 위의 5개년도에 그렇게 크게 나타나지 않는

다.

〈표 2-1〉 Kendall's Concordance Test 결과

연 도	출연자수	심사위원수	W	χ^2
1991	5	15	0.5182	16.7143 **
1990	5	20	0.3650	
1989	7	18	0.5013	
1988	6	21	0.6250	
1987	5	20	0.5415	
1986	4	21	0.2653	
(1985	2	21	1.0000)	
1984	4	20	0.2020	
1983	3	17	0.2630	
1982	4	19	0.4305	
(1981	2	22	0.4050)	
1980	5	19	0.4294	
1979	3	16	0.0820 *	
1978	3	15	0.2533	
1977	3	12	0.0069 *	
1976	3	15	0.6978	
1975	3	14	0.0051 *	
(1974	3	14	0.7347)	
1973	4	14	0.2061	
(1972	2	15	1.000)	
1971	3	15	0.1111 *	
1970	4	14	0.8286	
1969	3	11	0.2314 *	
1968	5	11	0.3901	
(1967	2	11	0.6694)	
1966	6	11	0.6874	

()안은 출연자 수가 2인인 경우.

*: 5%의 유의수준에서 각각수준치 이하인 계수.

**: N >7이고 k>20인 경우는 χ^2 분포로 근사.

N: 출연자수, k: 심사위원수.

이것은 피아노 부문에서 심사위원들간에 상당한 의견 차가 있었음을 의미한다. 1980년대 이후에는 새로운 심사위원의 구성비가 그 이전보다 높음에도 불구하고 이 검증방법에 의한 불일치의 경우가 나타나지 않는 것은 주목할만하다. 새로운 심사위원이 기존심사위원들과 비슷한 견해를 가지고 있어 심사위

원이 교체되어도 전체적으로 의견이 일치되어 있다고 볼 수도 있고 아니면 심사내용이 개선되었을 가능성도 보여준다.

일치계수검정방법은 전체적인 의견일치여부를 검정하는 것이나 개별심사위원 또는 심사위원 그룹간의 독립적 행동을 검정해 보는 것이 아니므로 별도의 분석이 필요하다.

3. Kendall 상관계수³⁾에 의한 독립성 검정

앞에서 제기된 바와 같이 각 심사위원이 독립적이기는 하나 임의적인 판단만으로 순위를 정하는 것은 공정성을 결여하는 결과를 콩쿠르 참가자에게 줄 수 있다.

심사위원판정에 대한 독립성검정이 앞 절에서 정의한 공정성을 평가하는 중요한 수단이 될 수 있다. Kendall 상관계수로서 독립성을 검정할 수 있는 경우는 본선참가자의 수가 4명 이상일 때이므로 해당 연도들은 1991, 1990, 1989, 1988, 1987, 1986, 1984, 1982, 1980, 1973, 1970, 1968, 1966 등의 13년간이다.

X = 특정 연도의 어느 특정 심사위원의 심사순위

Y = 특정 연도의 어느 특정 심사위원을 제외한 다른 모든 심사위원들의 심사순위의 종합심사순위

H_0 : X 와 Y 는 독립이다.

H_1 : X 와 Y 는 양의 상관관계를 가진다.

〈표 3-1〉은 귀무가설을 채택한 심사위원의 수와 비율을 연도별로 나타낸 표이다. 대회회수가 경과함에 따라 귀무가설을 기각하는 비율은 증가하는 경향을 띄지만 단정적이지는 않다. 1973년, 1982년, 1984년 등은 그 때 참가한 심사위원들의 상당수가 귀무가설을 채택하는 것을 알 수 있다. 1984년의 경우는 매우 특이한 경우로 모든 심사위원이 귀무가설을 채택하는 것으로 나타났다.

3) Spearman의 순위상관계수와 거의 동일한 결론을 얻을 수 있으며 소표본의 경우에 보다 우수하며 미지의 모수를 추정하는 것으로 볼 수 있다. Friedman의 검정은 각 관찰치가 확률적으로 관찰된 경우에 사용되며 순위에 동점이 많은 경우는 Goodman & Kruskal의 gamma 통계량을 사용해야 한다.

〈표 3-2〉는 켄달상관계수에 근거하여 귀무가설을 기각하지 못한 심사위원들이 개인이나 그룹으로 심사에 참가하지 않았다고 가정하면 심사결과가 달라지는 경우를 보여준다. 만약 최종심사결과가 달라진다면, 이들의 다른 심사위원들과 판이한 심사결과는 최종 순위결정에 지대한 영향을 끼쳤다고 결론지을 수 있다.

〈표 3-1〉 귀무가설 기각 심사위원의 비율

연 도	1991	1990	1989	1988	1987	1986	1984	1982	1980	1973
기각경우수	4	8	11	11	5	6	0	1	6	2
참가심사 위원수	15	20	18	21	20	21	20	19	18	14
비율(%)	26.7	40.0	61.1	52.4	25.0	29.6	0	5.3	33.3	14.3

분석결과 여섯 번의 경우는 어느 특정 심사위원이 심사에 참가하지 않았다고 가정하면 최종 심사결과가 달라졌다. Kendall 상관계수를 응용해 볼 수 있는 -참가자의 수가 4명이상에 해당하는- 13년 가운데 6년이 심사위원의 구성이 변하면 최종 심사결과가 변할 수 있었다는 것은 다른 심사위원과 다른 견해를 가진 심사위원들의 영향력이 지대했다고 말할 수 있다.

〈표 3-2〉 특정 심사위원이나 심사위원그룹이 심사에 참가하지 않았다고 가정했을 때 심사결과가 달라진 경우⁴⁾

연 도	특정 심사위원
1968	G1 F4
1973	.
1975 *	E2 F6 A6 D5
1977 *	D3
1982	.
1984	E6
1989	K1
1991	I3

4. 그룹화

독립성과는 별도로 그룹성과 공정성과의 관계를 분석해 볼 필요가 있다. 한

4) *로 표시된 연도는 본선진출자가 3인으로 Kendall 상관계수로 검정할 수 없는 해이다.

심사위원이 다른 심사위원과 꼭 같거나 비슷한 심사결과를 나타내는 것이 반드시 이상할 필요는 없다. 그러나 그 유사성의 정도가 크거나 빈도가 높을 경우에는 관심을 가질 필요가 있고 특정 인사들이 다년에 걸쳐 유사한 행동을 보인다면 극단적으로 담합의 가능성도 배제하기 어렵다. 이 경우는 독립성과는 반대되는 개념이지만 그룹적인 행동이 판정에 중요한 영향을 끼치게 되므로 주목해 보아야 한다. 극단적인 예는 그룹으로서 귀부가설을 기각하지 못하면서 판정에 결정적인 영향을 미치는 경우이다.

똑같은 심사결과를 보고한 심사위원들—콘테스트 본선진출자가 2인인 경우는 제외—을 찾아내 2, 3인이 여러번 그러한 형태의 심사를 한 경우 2, 3인 모두 심사위원으로 참여한 회수에 대한 상대도수를 구해보았다.

〈표 4-1〉에서 *표시된 그룹은 동일한 심사결과를 보여주긴 하지만 그 상대빈도가 0.1근방이므로, 담합적이거나 음약세계가 비슷하다고 결론짓기는 힘들다. **로 표시된 그룹들은 동일한 심사결과를 보여줄 가능성이 0.5 이상이어서 명시적으로나 묵시적으로 담합적인 개연성이 높고 적어도 상당히 유사한 평가기준을 가지고 있다고 보아야 할 것이다. 나머지 그룹들은 상대도수가 0.2에서 0.5 미만인데, 이들 그룹들도 그 정도는 약하지만 담합적인 행동을 할 개연성이 있다고 보아진다. A6-D5-F4와 D3-J5-K5 등은 세 심사위원들이 동일한 심사결과를 보고한 그룹들로 담합적 행동으로 판정할 여지가 있다.

이러한 담합적 행동의 개연성이 있는 심사위원들이 동시에 심사위원으로 위촉되었을 때 심사의 공정성에 문제가 제기될 수 있다. 왜냐하면, 이들의 심사결과가 최종심사결과에 결정적인 영향을 끼칠 수 있기 때문이다.

〈표 4-2〉는 각 심사위원이 자신과 같은 심사결과를 보고한 다른 심사위원들이 있는 경우의 수와 심사위원으로 참가한 회수에 대한 상대도수를 나타낸 표이다. 표를 해석하면 다음과 같다. 심사위원 D6는 심사위원으로 15회 참가하였는데 그 중 11회가 그룹에 속하는 행동으로 나타났다. 앞에 나왔던 표를 이용하여 설명하면 D6는 귀부가설을 기각하는 그룹행동에 속하는 경우가 대부분이고 독립적이면서 그룹에 속한 때는 1회이었다〈표 3-2〉. 또 이 행동은 다음에 나오는 경우와도 연결해서 생각해 볼 수 있다.

D6와 반대되는 경우가 A6와 E2이다. 이 두 위원은 그룹화된 행동을 하면서도 다수의 판정과는 다른 행동을 하는 경우가 상대적으로 많았고〈표 3-2〉 만약 심사에 참가하지 않았으면 최종결과를 변경시켰을 행동을 개별적으로 또한 그

〈표 4-1〉 2회 이상 동일한 심사결과의 빈도

그 룹	심사 참가(회)	동일한 결과(회)	상대숫수
A6-D5	15	4	0.267
D6-H4	9	4	0.444
F4-I5	12	4	0.333
A6-D5-F4	11	3	0.273
B6-H6	8	3	0.375
F5-G6	9	3	0.333
A3-F3	2	2	1.000 **
A6-C1	18	2	0.111 *
A6-D6	17	2	0.118 *
B6-H1	3	2	0.667 **
B6-H5	7	2	0.286
B6-I5	8	2	0.250
C1-G1	7	2	0.286
D3-J5-K5	4	2	0.500 **
D5-D6	17	2	0.118 *
D6-G5	8	2	0.250
E2-F5	9	2	0.222
F3-J5	4	2	0.500 **
F4-I6	9	2	0.222
G4-G5	5	2	0.400

롭적으로 하고 있다.

표현의 문제 때문에 생기는 오해가 있을 수도 있다. 같은 심사결과를 자주 나타낸 것은 답합의 여지가 있는 것으로 해석할 수도 있고 오히려 앞서 정의한 대로 더 공정한 심사위원으로 평가할 수도 있다. 다만 집단으로 표현된 동일성이 심사결과에 결정적인 영향을 미치는 경우는 심각한 문제로 받아들여져야 할 것이다. 그러므로 〈표4-2〉는 〈표3-2〉와 비교해서 평가해야 정확한 판단이 가능하다. G1, F4, D3는 〈표 3-2〉에서 볼 때 결정적 영향을 행사하였으나 〈표 4-2〉에서는 동일한 심사결과를 낸 동료심사위원들이 60% 이상의 경우에 있었던 것으로 계산되었다. 따라서 이들 심사위원들은 동일한 심사결과를 보인 다른 심사위원들과 함께 심사에 보다 강력한 영향을 끼친 경우이다. 반면에 E6는 동일한 심사결과를 보인 심사위원이 거의 없었는데 이와 같은 결정적 역할을 수행한 후에도 계속 심사위원으로 위촉되고 있어, 동아 콩쿠르가 추구하는 음악성에 비일관성을 야기했다고 볼 수 있다.

<표 4-2> 그룹에 속한 빈도와 참가회수에 대한 상대도수

본선진출자가 3인 이상

빈도	심사위원 (참가회수)	상대도수
11	D6(15)	0.733
10	B6(12) E2(15)	0.833 0.667
9	D3(14) D5(12)	0.643 0.750
8	A6(19) C1(15) F4(12)	0.421 0.533 0.667
7	F5(10) H4(8) I5(9)	0.700 0.875 0.778
6	G4(6) J5(9)	1.000 0.667
5	F1(7) F3(6) G5(7)	0.714 0.833 0.714
4	B1(5) H5(8) H6(4) I2(7) K5(5)	0.800 0.500 1.000 0.571 0.800
3	A3(3) B4(4) E4(4) F6(3)	1.000 0.750 0.750 1.000
	G1(5) G2(3) J3(3) J4(4)	0.600 1.000 1.000 0.750
2	C4(5) C5(5) D1(3)	0.400 0.400 0.667
	E1(5) H1(4) J1(6)	0.400 0.500 0.333

5. 최종심사결과와 같은 심사결과를 제출한 심사위원들

다음으로 고려한 것이 최종심사결과와 같은 판정을 한 위원들이다. 가치판단을 떠나서 또 담합여부를 떠나서 개별심사위원이 최종심사결과와 같은 판정을 하는 경우는 주목해 보아야 할 부분이다.

<표 5-1>은 최종심사결과와 같은 심사결과를 보고한 심사위원들을 본선진출자가 3인 이상인 경우에 나타낸 표이다. <표 5-2>는 <표 5-1>에 나타난 심사위원의 빈도와 심사에 참가한 회수에 대한 상대도수를 계산한 것이다.

10회 이상 위원으로 위촉받았으면서 개별심사결과가 최종판정과 같은 경우에서 가장 높은 빈도수는 0.385이며 D5가 이에 해당된다. D5는 그룹화행동을 하면서 최종결과에 가장 가까운 결과를 나타내는 행동을 하였으나 다른 위원들과의 일치성은 강하지 않았다<표 4-1>. <표 5-2>에서 보는 바와 같이 D5외에 상대도수가 30%를 넘는 위원은 10회 이상 참가의 경우 2인(D3와J5)이며 10회 이하 참가자의 경우에는 상대도수가 높은 편이나 물론 이 결과가 공정성을 직접 측정하는 지표가 될 수는 없으나 상대도수가 낮은 위원은 독립성이 높은 위원일 가능성이 많고 실제로 A6나 E2와 같은 심사위원들은 상대도수가 낮은 경향을 나타낸다.

F4의 경우는 12번 심사에 참가해 17%정도 최종심사결과와 동일한 심사결과

를 채점했으며 1968년도에는 결정적 역할을 수행하고 있음을 <표3-2>에서 볼 수 있다.

D5의 경우는 최종결과와 동일한 채점결과가 38%정도로 아주 높으나 1975년에는 결정적 역할을 수행하였다. D3는 31%정도로 동일한 채점결과를 보였으나 1977년에는 결정적 역할을 수행하였다. <표3-2>에서 결정적 역할을 수행한 G1, E6, K1, I3는 최종결과와 동일한 심사결과를 내는 경우는 거의 없는 것으로 조사되었다. 또한 1975년도 이외에 결정적 역할을 수행한 심사위원들은 독립적인 경우가 50%가 넘게 나타난다.

<표 5-1> 최종심사결과와 같은 순위로 심사한 심사위원들

연도	심사위원
1991	B4 E4 F3
1990	C6 I6 J1
1989	*
1988	G3 I6
1987	H5
1986	B6 D3 D6 G6 H1 K4
1985	**
1984	A6
1983	A6 B6 D5 E2 F5 H5
1982	K1
1981	**
1980	D6 G4 H4
1979	D3 D6 H4 K5
1978	D3 F4 G4 J5 K5
1977	D3
1976	A6 D4 D5 D6 F4 F5 I5 J4 J5
1975	A6 D5 E2 F6
1974	**
1973	A3 D5
1972	**
1971	B6 D5 D6 H6 I5 I6
1970	A3 B6 C1 C4 F3 H6 I6
1969	A6 C1 D3 G1 J5
1968	B6
1967	**
1966	*

*본선진출자는 3명 이상이었으나 해당심사위원이 없는 경우

**본선진출자가 2명인 경우

<표 5-2> 최종심사결과와 동일한 심사를 한 빈도수

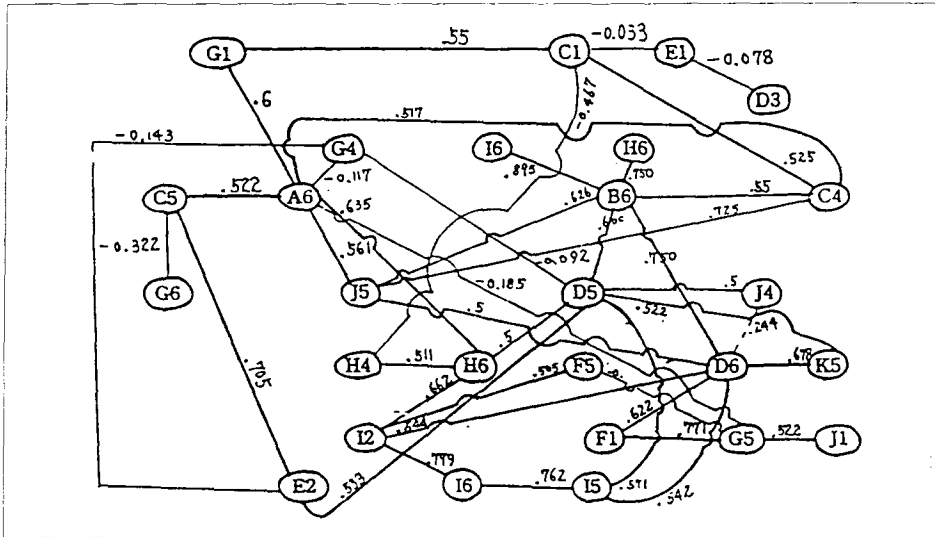
빈도	심사위원	상대도수
5	A6(19) D5(13)	0.263 0.385
4	D3(13) D6(15)	0.308 0.267
3	B6(12) C1(15) I6(12) J5(10)	0.250 0.200 0.250 0.300
	A3(3) E2(15) F3(5) F4(12)	0.667 0.133 0.400 0.167
2	F5(10) H4(8) H5(7) H6(6)	0.200 0.250 0.286 0.333
	I5(9) K5(5)	0.222 0.400

최종심사결과와 비슷한 순위로 심사한 위원의 수가 증가하거나 개별 심사위원이 심사에 참가할 때 최종심사결과와 비슷한 판정을 한 횟수가 올라갈수록 우연성이나 독립성에 의한 순위변동의 가능성은 줄어든다. 앞서의 여러 예에 자주 등장하는 A6와 E2의 경우 양 위원은 오랫동안 심사에 위촉되었으면서 동시에 개인 혹은 그룹으로서 심사에 결정적으로 다른 영향을 끼친 경우가 많지 않았음을 지적할 수 있다.

6. 평균적인 Kendall상관계수를 이용한 심사위원들간의 관계

<그림 1>은 연주자가 3인 이상이며 동시에 5회 이상 심사에 참가한 위원들의 개별 Kendall상관계수 중에서 상관계수평균이 0.5 이상이거나 負의 계수를 나타낸 그림이다. 5회 이상으로 한정된 것은 평균값의 안정성(신뢰성)을 높이기 위한 점이다. 즉 2회 동시참가의 경우보다는 5회 동시참가의 경우 Kendall상관계수의 평균이 보다 안정적이라고 할 수 있다. 또한 평균적인 Kendall상관계수의 계산에서 연주자가 3인 이상일 때 5회 이상 동시참가한 보다 많은 심사위원들을 포함하기 위한 것이다. 평균계수가 0.5 이상이라는 의미는 두 심사위원이 어느 정도 비슷한 판정을 하는 성향이 상대적으로 높다는 의미이고 한 사람을 사이에 둔 제3자와의 관계도 어느 정도 추정할 수 있다. 線分上에 수치가 표시되지 않고 연결되지 않은 부분은 평균적인 상관계수가 0.5 미만일 경우이다. 실제적인 의미는 상대적으로 독립적이라는 해석이 될 것이다. 0.5라는 수치가 任意的인것은 하지만 어떤 판단을 위한 기준으로는 유용할 수 있다. 만약 상관계수가 負로 나타나면 두 사람의 판정성향이 반대방향인 것을 의미하므로 관심의 대상이 된다.

〈그림 1〉 Kendall 상관계수를 이용한 심사위원들간의 관계



그림에서 5회 이상 심사에 참가한 26명 간에 개별 평균 상관계수가 보고되어 있다. 관심의 대상이 된 총 42개의 관계 중에서 8개가 負의 계수이며 그 중에는 계수의 크기가 -0.3 이상인 것도 있다. 앞에서 분석한 방법을 원용하여 이 그림이 의미하는 내용 중에서 몇 개의 예를 들어보면 다음과 같다. 위원번호 D5, J4, H6, B6들은 계수 0.5 이상의 관계로 연결된다. J4의 경우에는 심사에 지속적으로 참가하지 않은 이유 등 때문에 D5과만 관계가 나타난다.

D5와 B6는 여러 다른 위원들과 직접 혹은 간접적인 正의 관계를 맺고 있으며 D5는 13회 심사 중 2회의 귀무가설 기각실패, B6는 13회 심사 중 2회의 귀무가설 기각실패 경력이 있다.⁶¹⁾ B6와 D6는 서로 밀접한 관계가 있을 뿐 아니라 각각이 일정한 그룹을 형성하고 있는 상태이다. 다른 주요 위원들도 정도의 차이가 있지만 같은 현상을 보이고 있다. 이 결과를 조심스럽게는 하지만 조금 강하게 해석해 보면 G6와 J4는 상당히 혹은 매우 독립적인 성향을 가지고 있으나 J4의 경우 D5와 평균적으로 비슷한 경향을 나타내어 담합의 가능성도 있다고 볼 수 있다.

하나의 시나리오로써 B6심사위원과 0.5 이상의 평균적인 상관관계를 나타내는 심사위원이 다수 심사위원에 위촉되고 B6심사위원과 負의 상관관계나 상

6) <표 3-1>에서는 연주자가 4인 이상인 경우만 포함되었다.

관계수가 0.5 미만인 심사위원들이 소수로 심사위원으로 위촉되었을 경우를 고려하자. 이런 경우에는 B6위원의 음악성과 유사한 음악성을 가진 연주자가 상위에 입상할 것으로 예상된다. 따라서 심사위원의 구성은 콩쿠르의 최종결정에 중요한 영향을 끼치는 것으로 실증되었으며 이것은 공정성에 대한 비판의 대상이 될 수 있다.

다른 하나의 분석으로 A6와 E2는 거의 매년 심사위원으로 위촉되었으나 심사결과와 상관관계는 0.5 이하라는 것을 알 수 있다. 즉, 두 단골심사위원의 평가가 꽤 차이가 있음을 알 수 있다. 이 두 심사위원이 각각 심사에 참여하지 않았을 때 최종순위가 바뀐 해는 1975년인데 이 때는 다른 심사위원들도 결정적 역할을 하고 있다. A6의 경우 최종심사결과와 동일한 결과를 등위를 낸 경우가 26%정도이며 E2는 13%정도이다. 두 심사위원은 거의 매년 심사위원으로 위촉되었으나 최종심사등위와 다른 심사결과를 낸 경우가 대부분이었다. 다시 말해서 A6와 E2가 결과적으로 독립적으로 심사하고 있어서 콩쿠르에 참가하는 참가자들은 동아콩쿠르가 어떤 음악세계를 추구하는가를 판단하기 힘들다고 평가할 수도 있다.

IV. 요약과 결론

동아콩쿠르는 심사위원별 심사결과를 공개하므로 투표에 의한 순위결정이나 당락결정과정을 분석할 근거는 있다. 다른 대부분의 심사들-예를 들면 예술이나 문학경연 또는 각종 고시-은 내용이 공개되지 않은 만큼 더 많은 문제를 가질 수 있다. 순위투표 모형과정에서 제일 문제시되는 것은 공정성과 일관성이다. 이 두 개념이 어떻게 정의되어야 하는 것 또한 문제가 되지만 이 부분은 분석을 통하여 개선책을 기대할 수 있다.

다른 하나의 논점은 투표결과가 우연성에 의해 영향력을 어느 정도 받느냐의 문제이다. 만약 심사결과가 심사위원선정에 따라 달라진다면 콩쿠르의 존립목적을 다시 생각해 보아야 한다. 이 원리는 고시나 다른 경우에도 마찬가지로 고려되어야 한다.

본 연구에서 나타난 결과와 그 의미를 다음의 몇 가지로 정리해 볼 수 있다.

1) 원칙적으로 심사위원은 자주 교체되는 것이 교체가 드문 경우보다 결과를 개선시킬 수 있다. 그러나 만약 교체되는 심사위원들이 옛날보다 더 많은 독립

성을 나타내면 오히려 改惡의 결과를 초래할 수도 있다. 본 연구결과 교체율이 높을수록 독립적인 심사결과가 나타난다는 증거는 없다. 특히 독립적인 행동은 새로 교체되어 들어오는 위원이 아니라 계속 위촉되는 경우에 더 많이 나타났다. 1990년과 1991년의 경우 매년 5명의 심사위원이 독립적인 행동을 하였는데 그 해 처음으로 참가한 위원은 1991년의 2명뿐이고 나머지는 3회에서 24회까지 계속 위촉된 위원이었다.

2) 개별심사위원의 행동에 관계없이 심사결과는 심사위원간 일치하는 방향으로 나타나야 일관성과 신뢰성을 보여 줄 수 있다. 동아콩쿠르의 경우 1980년 이전에는 심사위원간 의견불일치가 자주 나타났고 그 이후에는 전혀 나타나지 않아서 심사위원구성이나 심사내용이 개선된 것으로 볼 수 있다.

3) 주최기관이 심사위원의 개별 행동에 대해 분석하지 않으면 공정성과 일관성에 문제가 있을 수밖에 없다. 본 연구결과에 의하면 독립적으로 행동하거나 순위에 결정적 영향을 행사한 위원이 거의 매년 나타났다. 1)과 같은 경우 순위가 뒤바뀌는 것은 어떤 경우에도 피하는 것이 좋으므로 심사위원 선정에 세심한 분석이 필요하다. 고시는 동아콩쿠르보다 더 심각한 문제를 내포하고 있을 가능성이 있다.

4) 비단 음악부문 뿐만 아니라 다른 부분에도 이 원칙은 적용되어야 한다. 주최기관이나 평가기관이 확실한 목표를 분명히 정하고 그에 따른 위원선정에 노력해야 한다. 고시와 같은 경우는 본 연구와는 다른 방법으로 공정성과 일관성을 평가해 볼 수 있다.

5) 심사위원으로 자주 위촉되는 사람들은 아마도 有名度나 업적에서 더 뛰어난 경우라고 유추할 수 있다. 그러나 자주 위촉되는 사람은 어떤 형태로든지 문제를 일으킬 소지가 많은 것으로 나타났다. 자주 위촉되는 위원은 대체적으로 그룹화되는 성향을 나타내었고 또 그룹간은 서로 다른 性向을 나타내는 경향이 나타났다. 이것은 좋은 의미에서 음악성이나 음악관을 같이 하는 뜻도 있으나 경제학적인 의미의 암묵적 담합의 결과도 초래할 위험이 있다는 것을 뜻한다.

참 고 문 헌

- 김성숙, 一般化 可能度 理論을 이용한 教師行爲 觀察에 있어서 誤差源分析, 교육평가 연구, 제 3권, 제 1호, 1989.
- 송문섭, 박창순, 비모수통계학개론, 자유아카데미, 1989.
- 차영준, SPSS /PC+ 를 이용한 비모수통계학, 자유아카데미, 1993.
- M. G. Kendall, *Rank Correlation Methods*, Chareles & Griffin Co. Ltd 1970.
- Robert F. Fagot, Reliability of Rating for Multiple Judges: Intraclass Correlation and Metric Scales, *Applied Psychological Measurement*. Vol, 15, NO. 1, 1991.
- Douglas E. Critchlow & Joseph S. Verducci, An Omnibus Test for Systematic Changes in Judges' Rankings, *Journal of Educational Statistics*, Spring, Vol. 17, NO. 1, 1992.