

SFS(Sequential Forward Selection) 알고리즘에 의해 선정된 특징들을 이용한 교통 사운드정보의 내용 기반 검색

Content-Based Traffic Sound Information Retrieval
using Selection Features by Sequential Forward Selection algorithm

정윤주

(광운대학교 컴퓨터공학과 석사과정)

최기호

(광운대학교 컴퓨터공학과 교수)

Key Words : 교통정보, 내용기반 사운드검색, 특징선정

목 차

- I. 서론
 - II. 선정된 특징열에 의한 내용 기반 검색
 - 1. 사운드 특징 추출
 - 2. K-means 에 의한 특징별 분류
 - 3. SFS 에 의한 특징열 선정
 - III. 실험결과 및 분석
 - IV. 결론 및 향후계획
- 참고문헌

I. 서론

과거 경제 부흥 국가라는 이름에 걸맞은 급속한 경제성장은 수송관련 기반시설의 현대화와 아울러 예상치 못했던 수송수요를 창출했으며, 결국 이러한 수송수요는 교통 혼잡을 야기시키기 시작했다. 이에 미래사회에 부흥하는 교통 혼잡과 사고의 감소를 위해 주요도로 정체구간, 교차로, 또는 교통사고가 빈번한 위치에 신호위반 단속시스템과 감시카메라를 설치하여 현재 교통상황을 파악해 원활한 도로교통 환경정보를 제공하거나 교통사고와 뺑소니 등의 범죄 예방 및 증거자료로 활용하고 있다.

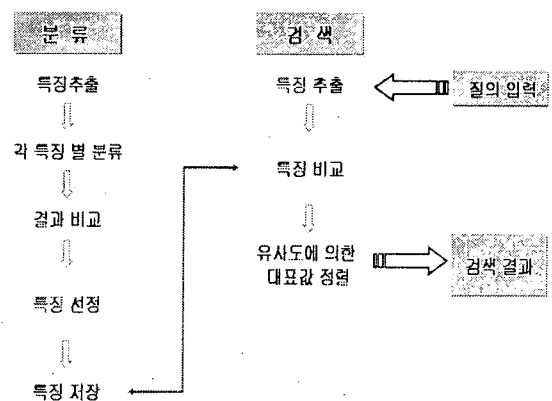
그러나 도로 교통 상황을 감시하고 교통 정보를 추출하는 현재의 기술은 영상에 많이 의존하고 있는 실정으로 영상정보에 비해 상대적으로 적은 용량을 차지하는 사운드 정보를 함께 사용한다면 보다 더욱 정확한 정보수집과 분류가 가능할 것이다. 예를들어 자동차의 충돌소리, 급정거소리 등의 특정소리가 감지되면 사고로 인지하고 전후 수 초간을 저장함으로써 저장용량의 감소를 돕는 역할을 할 수도 있다.

이렇게 수집된 사운드 정보에 대한 분류는 기존의 수작업으로 이루어지던 분류시스템과 달리 사운드 정보의 내용을 수학적으로 분석하여 구조화된 기준에 따른 대표적인 특징을 추출하고 컴퓨터를 통해 데이터를 체계적인 구조로 색인화하고 검색한다.

사운드 정보를 검색하기 위한 방법으로는 크게 예제질의 방식(QBE : Query By Example)과 허밍질의방식(QBH : Query By Humming)으로 구별되는데 모든 파일 포맷에 적용가능한 예제 질의 방식이 주로 사용되고 있다. 예제 질의 방식은 디지

털 신호처리 기법을 사용해 사운드의 물리적 특징 및 지각적 특징을 추출한 후 DB내의 사운드와 비교 검색하는 방법이다.

기존 내용기반 오디오 및 사운드 정보 검색에 관련된 연구는 1997년 미국 Audible Magic 사의 MuscleFish[1]로 시작하여 최근에는 ACM ISMIR 학술회를 중심으로 다양한 연구 논문이 발표되고 있고 특히 다양한 정보를 갖는 사운드의 공통적인 특징을 찾아내기 위한 연구는 국내외적으로 활발히 진행 중이다.



<그림 1> 선정된 특징열을 사용한 사운드 내용기반 검색

대표적인 예제 질의 방식으로 E. Wold[1]는 MuscleFish를 통해 15초 미만의 동물소리, 기계소리, 악기소리, 음성 등의 사운드 효과들에 대하여 신호의 크기 (loudness), pitch, 밝음 (brightness), 대역폭, 하모니 등의 특징들을 추출하여 DB내에서 유사한 오디오를 검색하였으며 T. Chang[2]는 HMM을 이

용한 계층적 구조의 오디오 분류 및 검색 시스템을 제안하였고, 음성, 박수 소리, 폭발음, 새소리 등의 환경소리를 대상으로 하여 약 80%의 성공률을 보였다.

본 논문에서는 내용기반 사운드 검색에 사용되는 많은 특징들중 Sequential Forward Selection 알고리즘을 이용하여 선정된 최적의 특징열을 사용해 연산량을 감소시키고자 하였으며 구성은 다음과 같다. 1장 서론에 이어 2장에서는 사운드에서 추출된 각각의 특징에 대해 분류작업을 수행 후 SFS 알고리즘을 적용시켜 적절한 특징열을 선정하고, 3장에서 실험결과를 분석하며 4장에서 결론과 향후과제를 마지막으로 논문을 맺으며 간략한 순서도는 <그림 1>과 같다.

II. 선정된 특징열에 의한 내용기반 검색

1. 사운드 특징 추출

특징추출이란 사운드의 특성을 반영하여 분석하고 이를 이용하여 분류나 검색에 필요한 샘플들을 추출하는 것을 말한다. 사운드의 신호 샘플들은 고정된 사이즈의 창 크기(Window duration) 단위로 분석이 이루어지는데 Rectangle window에 비해 주파수 영역에서의 모델링에 더 정확한 역할을 수행할 수 있는 Hamming window가 주로 사용된다. 이 특징들은 측정가능한 특성을 기반으로한 통계적이고 수학적인 물리적 특징, 사람에게 의해 인지되어 보고되어 지는 특성을 기반으로 한 인지적특성과 계수영역특징 등으로 나뉠 수 있다.

1) Energy

물리적 특징중 가장 수월하게 측정 가능한 것중 하나인 energy는 어떤 시간 내 신호의 양을 측정하는 것으로 대개 윈도우를 사용한 short-time을 기본으로 사용된다. 에너지는 동적인 범위뿐 아니라 silence부분을 발견하는데 사용되며 음성 음 세그먼트와 무성음 세그먼트의 식별에도 중요한 특징이 된다.

2) ZCR

ZCR(Zero Crossing-Rate)의 값은 energy와 함께 실제 사운드 영역을 검출하는 역할 외에 스펙트럼에서 에너지가 집중되는 주파수를 찾는 데 유용한 특징 파라미터로 널리 사용되는 특징이다. 이는 분석구간 프레임 내에서 신호파형이 영점(zero)측과 교차하는 횟수를 의미하는데 무성음의 경우 진폭이 크지 않고 불규칙적인 진동이 계속 있으므로 일정한 만큼만 중심축을 통과하는 유성음보다 영교차율이 크다. 또한 ZCR은 음성신호와 음악의 분류에도 유용하게 사용되는데 음성신호와 비교할 때 음악의 ZCR 파형은 눈에 띄게 낮은 분산과 평균 진폭을 갖는다.

3) Spectral Features

신호의 스펙트럼은 신호내 주파수의 분포를 기술하고 있는 신호의 time-varying 스펙트럼인 스펙트로그램을 발생시킨다. 이를 위해 신호는 STFT(Short Time Fourier Transform)를

통해 프레임 안으로 쪼개어지게 되고 각 프레임 내에서 계산되어진 time-varying 스펙트럼을 통해 신호변화의 주파수내용을 측정해 Spectral Centroid, Spectral Rolloff, Spectral Flux 등을 산출해 분류를 위해 사용하고 있다.

우선 Spectral Centroid에서 centroid란 스펙트럼의 중심을 뜻하는데 스펙트럼의 형태를 측정하는 방법 중의 하나이다. Spectral Rolloff는 스펙트럼 magnitude 분포의 80%가 집중되어 있는 주파수 이하를 나타내는데 rolloff는 스펙트럼의 형태와 낮은 주파수 영역에 신호의 에너지가 얼마나 집중되어 있는지를 보여준다. 마지막으로 Spectral Flux는 연속된 스펙트럼 분포에서 정규화 된 magnitude들 간의 차이를 제공해서 구할 수 있는데 정규화 된 magnitude인 Flux는 스펙트럼 변화의 양을 계산할 수 있다.

4) Perceptual Features

사람에 의해 인지되어지는 perceptual 특징에는 loudness, pitch, brightness, bandwidth, harmonicity, timbre, rhythm 등이 포함된다.

timbre은 같은 높이의 음과 같은 크기를 갖는 사운드라 하더라도 다른 악기나 음성에 의해 만들어졌다면 발음체의 차이나 진동방법에 따라 음의 성분차이에서 생기는 감각적인 특성, 즉 음색을 의미한다. 예컨대, 음의 높이는 같아도 피아노와 바이올린의 음색이 전혀 다른 느낌을 주는 것은 바로 이와 같은 이유 때문이다.

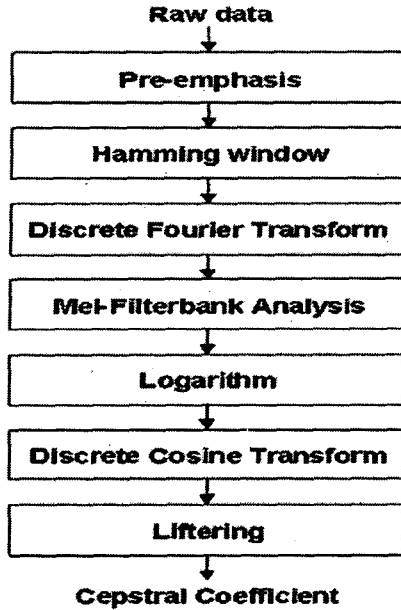
두 번째로 loudness는 사람의 청각이 소리의 물리적인 세기가 같더라도 주파수에 따라 감각량에 차이가 난다는 점과 그 감동의 차가 세기의 절대값에 의해서도 변화한다는 특성을 실험적으로 구한 척도를 의미한다.

세 번째 Pitch는 매우 중요한 인지적 특징중 하나로 사운드에 관련된 많은 정보를 표현하고 있으며 마지막으로 리듬은 규칙적으로 일정 시간에 걸쳐 발생하는 특성을 갖고있는 특징으로 원래 음성신호와 음악을 위해 정의된 부분이지만 footstep, clock tick, telegraph machine 등의 사운드 클립에서 패턴의 변화를 표현함으로써 매우 중요한 역할을 담당하기도 한다.

5) LPC[3]

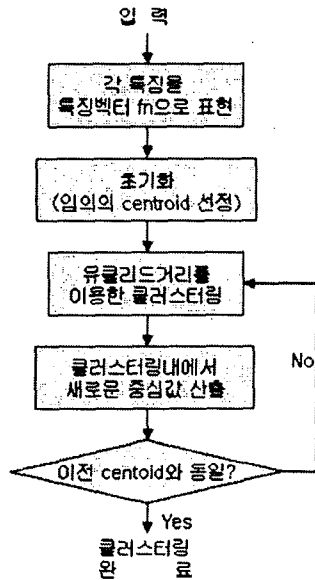
LPC(Linear Predictive Coding)는 인간의 발성 모델에 입각해서 음성 신호를 부호화하는 방법으로 과거의 일정 개수의 샘플 값들에 계수를 각각 곱하고 이를 총 합한 값으로 현재의 샘플 값을 예측하려는 시도에서 출발한다. 또한 LPC를 기반으로 음성 인식에 효과적인 켈프스트럼 계수로 변환하여 사용하기도 하는데 켈프스트럼은 로그 스펙트럼을 푸리에 변환한 것으로 성도 모델에 관한 파라미터와 입력 기저신호에 관한 모델을 선형적으로 분리함으로써 성도 모델을 나타내는 파라미터 추출에 유리하게 사용된다.

6) MFCC



<그림 2> MFCC 특징 추출

MFCC(Mel Frequency Cepstrum Coefficient)[4]는 사람의 귀가 주파수 변화에 반응하게 되는 양상이 선형적이지 않고 로그 스케일과 비슷한 멜 스케일을 따르는 청각적 특성을 반영한 계수 추출 방법이다. 멜 스케일에 따르면 낮은 주파수에서는 작은 변화에도 민감하게 반응하지만, 높은 주파수로 갈수록 민감도가 작아지므로 특징 추출 시에 주파수 분석 빈도를 이와 같은 특성에 맞추는 방식이다[5]. MFCC의 추출방법은 <그림 2>와 같다.



<그림 3> K-means 클러스터링

2. K-means 에 의한 특징별 분류

클러스터링이란 n개의 객체를 대상으로 p개의 특징들을 측정하였을때, 이 변수 p를 이용하여 n개의 객체들 사이의 유사성의 정도를 측정하여 객체들을 가까운 순서대로 연결해 나가는 분석방법을 말한다. 클러스터링은 최종적으로 모든 객체가

하나의 클러스터링이 되는 계층적 클러스터링방법과 전체 객체를 특정 클러스터수로 최적 분배하는 방법인 비계층적 클러스터링방법으로 나뉘는데 본 논문에서는 다른 클러스터링에 비해 대량의 데이터를 클러스터링하는데 유용하게 사용되는 알고리즘인 K-means 클러스터링을 사용하였다. 사용된 K-means 클러스터링의 알고리즘이 <그림 3>에 나타나있다.

K개 만큼의 초기 centroid가 정해졌으면 각 프레임에서 추출된 특징벡터별로 클러스터링을 수행한다. 특징벡터 f_n 과 각각의 centroid의 유클리드거리를 측정후 가장 가깝게 위치한 centroid에 해당하는 클러스터로 배정한다. 이때 클러스터가 형성되는 즉시 새로운 centroid를 산출하면서 클러스터링을 반복해나간다.

$$C_r(i+1) = \sum_{n=0}^N \sum_{a=0}^A \{pixn \mid \|f_n - P_r(i)\| \leq \|f_n - P_a(i)\|\} \dots\dots\dots(1)$$

$$P_r(i+1) = \frac{1}{size[C_r(i+1)]} \sum_{n \in C_r(i+1)} f_n \dots\dots\dots(2)$$

식(1)에서 C_r 은 r번째 클러스터, i는 클러스터링의 반복횟수를 의미하며 N과 A는 각각 총특징 수와 전체클러스터의 개수를 의미한다. 식(2)의 size는 클러스터내의 특징수를 의미하며 centroid를 위한 평균값산출을 위한 식이다.

3. SFS 알고리즘에 의한 특징열 선정

1971년 A.W.Whitney에 의해 소개된 SFS(Sequential Forward Selection)[6]는 최소한의 특징열로 구성된 서브셋으로 가장 좋은 효과를 얻기위한 알고리즘으로 구현이 간단해 광범위하게 사용되고 있다. <표 1>은 SFS의 간단한 알고리즘을 나타낸다.

1. 비어있는 서브셋으로 시작 $S_0 = \{\emptyset\}$
2. 서브셋에 미포함된 특징중 최상의 효과를 보인 특징을 선정 $x' = \underset{x \in X}{\operatorname{argmax}} J(S_n + x)$
3. 서브셋 갱신 $S_{n+1} = S_n + x'$; $n = n+1$
4. 2단계 반복수행

<표 1> Sequential Forward Selection 알고리즘[7]

초기에 공집합으로 시작해 첫단계에서 모든 특징중 가장 높은 정확성을 갖는 하나의 특징을 선정하여 분류작업을 수행하고 다음단계에서는 두 번째 정확성을 갖는 특징이 추가되어 두개의 특징을 조합해 분류작업이 수행되어진다. 이렇게 각 단계에서 특징들을 하나씩 추가하는 작업을 반복해나간다면 가장 좋은 성공률을 나타내는 특징부터 순차적으로 나열되어진 새로운 특징열이 선정되어지고 이 특징열을 이용한 분류작업의 정확성은 어느 한 점에서 더 이상 크게 향상되지 않게 된다.

4. 선정된 특징열을 이용한 사운드 검색

본 논문에서 제안한 내용기반 사운드 정보 검색에서는 예

제 질의 방식을 사용하게 된다. 질의데이터에서 SFS 알고리즘에 의해 선정된 특징열을 추출후 분류된 대표값들과 비교하며 이때 비교는 유클리드 거리를 사용한다.

III. 실험결과 및 분석

사용 특징	카테고리별 평균 분류 성공률(%)						총분류 성공률(%)	순위
	시동	도어	경적	충돌	주행	급정거		
ZCR	41.2	38.4	35.6	40.0	51.5	37.2	40.65	10
Energy	42.8	43.1	42.7	43.0	51.0	42.3	44.15	8
Centroid	42.6	41.7	43.0	39.5	50.6	41.1	43.08	9
Rolloff	26.4	37.2	32.0	36.4	28.0	41.2	33.53	11
Flux	43.1	44.6	44.2	43.5	51.4	43.2	45.00	7
MFCC 1	7.2	7.2	11.3	12.1	9.4	8.2	9.23	14
MFCC 2	52.6	46.3	54.2	54.7	40.4	52.8	50.17	1
MFCC 3	52.4	46.1	54.3	54.4	40.4	52.1	49.95	2
MFCC 4	51.7	45.2	54	53.8	40.1	51.6	49.40	3
MFCC 5	51.1	44.9	53.9	53	40.2	50.8	48.98	4
MFCC 6	49.0	42.6	43.8	48.2	40.9	46.4	45.15	6
MFCC 7	7.4	7.2	11.0	12.5	10.0	8.6	9.45	12
MFCC 8	49.8	43.2	52.1	51.6	38.1	49.2	47.33	5
MFCC 9	7.2	5.7	10.9	8.1	9.2	4.9	7.67	17
MFCC 10	7.4	7.2	11.6	12.0	10.2	8.0	9.40	13
MFCC 11	7.0	6.4	10.5	9.5	9.1	7.2	8.28	15
MFCC 12	6.1	6.4	10.8	9.2	8.4	5.6	7.75	16

<표 2> 특징별 분류 수행결과

본 논문에서 제안한 사운드 내용기반 검색의 실험을 위하여 도로내에서 들을 수 있는 여러 가지 사운드를 대상으로 실험하였다.

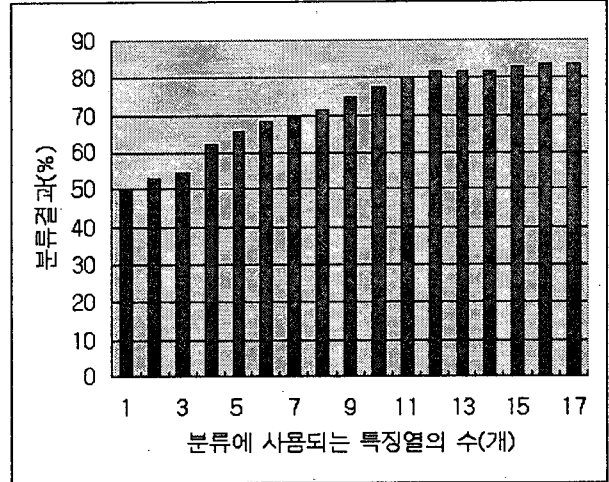
실험에 사용된 500개의 사운드는 44.1Khz, 16bit, mono로 샘플링 되었으며 사운드 신호는 시간에 따른 신호의 주파수 성분이 일정한 특성을 나타내도록 23ms 크기의 윈도우를 사용해 프레임단위로 나눈후 각각의 프레임에 hamming window를 씌웠다. 500개의 데이터는 차문여닫는소리, 시동소리, 주행소리, 급정거소리, 부딪히는 소리, 경적소리 등으로 이루어져있으며 이중 80%를 분류데이터로 사용하고 20%를 테스트데이터로 사용하였고 제한된 DB에서의 정확성을 위해 이를 랜덤하게 하여 100회 반복하여 실험하였다.

사용한 특징은 물리적 특징인 energy, ZCR과 spectral 특징인 spectral centroid, spectral rolloff, spectral flux와 계수특징인 MFCC 이며 <표 2>는 클러스터링을 사용해 특징별 분류를 수행한 결과를 나타낸다.

재정렬 결과			
순위	특징	순위	특징
1	MFCC_2	10	ZCR
2	MFCC_3	11	Rolloff
3	MFCC_4	12	MFCC_7
4	MFCC_5	13	MFCC_10
5	MFCC_8	14	MFCC_1
6	MFCC_6	15	MFCC_11
7	Flux	16	MFCC_12
8	Energy	17	MFCC_9
9	Centroid		

<표 3> 분류결과의 순위에 따른 재정렬

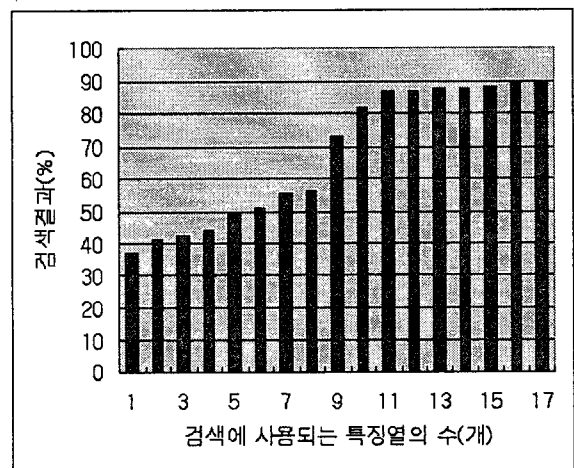
카테고리별 분류결과는 각 카테고리내의 총데이터중 알맞게 분류된 결과를 표현했으며 총 분류결과는 각 카테고리별 분류결과를 평균한 결과를 나타내고 그 결과에 따른 순위를 표현했으며 이를 재정렬해 <표 3>에 나타내었다.



<그림 4> 특징의 순차적 조합을 사용한 분류 결과의 비교

<그림 4>는 재정렬된 특징들을 순서대로 조합해 조합에 따른 분류결과를 비교한 그래프이다. 그림에서 볼 수 있듯이 12개의 특징을 조합한 결과부터는 특징의 증가가 분류결과의 향상에 크게 영향을 미치지 않기 때문에 선정된 최소의 특징열인 (MFCC 2차, 3차, 4차, 5차, 8차, 6차, spectral flux, energy, spectral centroid, ZCR, spectral rolloff)를 사용하여 분류한 DB와 질의데이터를 사용해 내용기반 검색을 수행하면 감소된 연산량으로 전체 특징을 사용했을때와 유사한 효과를 얻을 수 있다.

<그림 5>에서 각각의 특징열 조합에 따른 검색결과를 확인할 수 있다.



<그림 5> 특징의 순차적 조합을 사용한 검색 결과의 비교

IV. 결론 및 향후계획

본 논문은 현재 교통상황을 표현하고 있는 사운드 정보의

효과적인 분류와 검색을 위해 SFS(Sequential Forward Selection) 알고리즘을 이용하여 좋은 결과를 나타낼 수 있는 특징을 선정하는 방법을 제안하고 있다.

영상정보에 비해 상대적으로 적은 용량을 차지하는 사운드 정보의 이용은 영상정보만을 사용한 정보수집·분류에 비해 더욱 정확한 분류와 검색을 가능하게 한다. 이는 복잡한 도로의 교통소통을 원활하게 하거나 사고에 대한 증거수집과 예방을 하는 감시카메라 등에 적용시킬 수 있으며 선정된 특징을 사용한 사운드 분류작업은 연산량을 감소시켜 빠른 분류를 가능하게 한다.

향후 연구에서는 복잡한 배경음이 섞인 데이터에 적용시킬 강력한 잡음제거에 관한 연구가 필요하며 좀더 다양한 데이터에 대한 분류와 검색시스템의 구현이 필요하다.

참고문헌

1. E. Wold, T. Blum, D. Keislar, and J. Wheaton, "Content-based classification, search and retrieval of audio," *IEEE Multimedia*, 3(2), 1996
2. T. Zhang and C. -C. Jay Kuo, "Hierarchical System for Content-based Audio Classification and Retrieval," *Proceedings of SPIE's Conference on Multimedia Storage and Archiving Systems III*, SPIE, 3527, pp. 398-409, Boston, Nov., 1998.
3. J. Makhoul, "Linear prediction: A tutorial overview", *Proceedings of the IEEE*, Apr. 1975
4. B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling", In *Proc. Int. Symposium on Music Information Retrieval (ISMIR)*, 2000
5. M. Slaney, "A critique of pure audition", *Computational Auditory Scene Analysis*, 1997
6. A.W. Whitney. A direct method of nonparametric measurement selection. *IEEE Trans. Comput.*, 20:1100-1103, 1971.
7. http://research.cs.tamu.edu/prism/lectures/pr/pr_111.pdf