

Unifying Automated Fractionated Atrial Electrogram Classification using Electroanatomical Mapping Systems in Persistent Atrial Fibrillation Studies

Tiago P Almeida¹, Gavin S Chu^{2,3}, João L Salinet^{1,4}, Frederique J Vanheusden¹, Xin Li¹, Jiun H Tuan³, Peter J Stafford³, G André Ng^{2,3,5}, Fernando S Schlindwein^{1,5}

¹Department of Engineering, University of Leicester, UK

²Department of Cardiovascular Science, University of Leicester, UK

³University Hospitals of Leicester NHS Trust, UK

⁴Engineering, Modelling and Applied Social Sciences Centre, Federal ABC University, Brazil

⁵National Institute for Health Research Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, UK

Abstract

Ablation targeting complex fractionated atrial electrograms (CFAE) for treating persistent atrial fibrillation (persAF) has shown conflicting results. Differences in automated algorithms embedded in NavX (St Jude Medical) and CARTO (Biosense Webster) could influence CFAE target identification for ablation, potentially affecting ablation outcomes. To evaluate this effect, automated CFAE classification performed by NavX and CARTO on the same bipolar electrograms from 18 persAF patients undergoing ablation was compared. Using the default thresholds, NavX classified 69±5% of the electrograms as CFAEs, while CARTO detected 35±5% (Cohen's kappa $\kappa \approx 0.3$, $P < 0.0001$). Both primary and complementary metrics for each system were optimized to balance CFAE detection for both systems. Using revised thresholds found from receiver operating characteristic curves, NavX classified 45±4%, while CARTO detected 42±5% ($\kappa \approx 0.5$, $P < 0.0001$). Our work takes a first step towards the optimization of CFAE detection between NavX and CARTO by providing revised thresholds to reduce differences in CFAE classification. This would facilitate direct comparisons of persAF CFAE-guided ablation outcome guided by NavX or CARTO.

1. Introduction

Atrial fibrillation (AF) is the most common sustained arrhythmia in clinical practice and a leading cause of hospitalization and death [1]. Although radiofrequency catheter ablation has been consolidated as the most accepted percutaneous procedure for AF treatment [1, 2], ablation is still suboptimal in patients with long-term or persistent atrial fibrillation (persAF) due to insufficient understand-

ing on the mechanistic interaction between relevant atrial substrate and the initiation and maintenance of AF. Atrial electrograms (AEGs) with fractionated activity are believed to represent AF substrate [3]. The ablation of atrial substrate hosting complex fractionated atrial electrograms (CFAEs) has been accepted by many as a useful additional therapy for persAF treatment [3-6]. Disparities in CFAE-guided ablation outcomes have, however, cast doubt on the efficacy of this approach, with reported success rate varying from 14% to 95% [3-6]. Automated CFAE detection can be performed by algorithms embedded in commercial 3-dimensional (3D) electroanatomical mapping systems [6, 7]. These algorithms incorporate CFAE characteristics as initially described by Nademanee *et al.* [3]. Each algorithm, however, considers different premises to quantify fractionation in the modified atrial substrate. The two commercial systems most frequently used in clinical practice are the NavX™ (St. Jude Medical, St. Paul, Minnesota) [6] and the CARTO (Biosense Webster, Diamond Bar, California) [7]. Those systems provide primary metrics to assess CFAE objectively [NavX: CFE-Mean; CARTO: Interval Confidence Level (ICL)], and complementary metrics to further support the electrophysiology procedure [NavX: CFE-StdDev; CARTO: Average Complex Interval (ACI), Shortest Complex Interval (SCI)]. There are no defined thresholds for the complementary metrics to characterize CFAEs.

Differences in automated algorithms embedded in commercial systems might influence CFAE target identification for ablation, potentially affecting ablation outcomes [8, 9]. In this study, we report a direct, quantitative comparison of automated CFAE classification performed by algorithms embedded in NavX and CARTO. We investigated CFAE areas where both systems agree in order to customize their parameters and improve their mutual sensitivity and specificity for CFAE classification.

2. Methods

2.1. Electrophysiological study

The study population consisted of 18 persAF patients (16 male; mean age 56.1 ± 9.3 years; history of AF 67.2 ± 45.6 months) referred to our institution for first time catheter ablation [10]. Study approval was obtained from the local ethics committee and all procedures were performed with full informed consent. 3D left atrial (LA) geometry was created using Ensite NavX. Pulmonary vein (PV) isolation was performed followed by the creation of linear roof lines – defined as contiguous anatomical line – using a deflectable, variable loop circular PV mapping catheter (Inquiry Optima, St. Jude Medical). External electrical cardioversion was used to restore sinus rhythm if this was not achieved during the ablation procedure. No additional ablation targeting CFAE was performed in this study. Sequential point-by-point bipolar AEGs were collected from 15 pre-determined atrial regions before and after LA ablation for each patient [10]. A total of 797 AEGs were recorded from the LA (sampling frequency 1.2 kHz), and band-pass filtered within 30–300 Hz.

2.2. CFAE classification optimization

Each AEG (2.5-s), with their corresponding CFE-Mean and CFE-StdDev, were exported from NavX. A validated (100% agreement) offline MATLAB algorithm was used to compute ICL, ACI and SCI for CFAE identification according to CARTO criteria. CFE-Mean and CFE-StdDev were measured using the standard NavX settings (30-120 ms), and ICL, ACI and SCI were measured using the standard CARTO settings (50-110 ms) [6, 7]. $ICL \geq 7$ was used as the default threshold for CARTO CFAE categorization [4].

CFAE classification was performed on 697 randomly sampled AEGs (out of the total 797), first using CFE-Mean and then ICL. This dataset was used to create receiver operating characteristic (ROC) curves and hence obtain the optimum sensitivity and specificity thresholds for both metrics to counterbalance CFAE classification, using the counterpart metric as the comparator [11]. More specifically, the CFAE classification (CFAE / non-CFAE) according to the $CFE-Mean \leq 120$ ms [6] was used to create a ROC curve by varying ICL. The optimum threshold for ICL was identified based on the optimum sensitivity and specificity on the ROC curve – defined as the point on the curve with the shortest distance to the top left corner of the graph. Similarly, the CFAE classification (CFAE / non-CFAE) according to the $ICL \geq 7$ [4] was used to create a ROC curve by varying CFE-Mean. The optimum threshold for CFE-Mean was identified based on the optimum sensitivity and specificity on the ROC curve. Area under the ROC curve (AUROC) and

the P-value were also calculated. This process was iterated thirty times, each time with a different dataset of randomly sampled AEGs for ROC curve construction (697 AEGs), giving a total of thirty ROC curves for ICL and thirty for CFE-Mean in order to minimize data sample and selection biasing.

The revised thresholds for both CFE-Mean and ICL found in the ROC curves were used to perform a new CFAE classification on the thirty sets of 697 randomly sampled AEGs. In this ‘new’ classification, an AEG was classified as CFAE only if both CFE-Mean and ICL agreed with the classification using their revised thresholds. These classifications were used to create ROC curves and hence obtain the optimum sensitivity and specificity thresholds for the complementary metrics – CFE-StdDev, ACI and SCI.

The revised thresholds found from the ROC curves for both NavX – CFE-Mean and CFE-StdDev – and CARTO – ICL, ACI and SCI – were validated using the remaining 100 AEGs (thirty sets of 100 AEGs). For each of the thirty datasets, CFAE classification was performed using the combined assessment of both primary and complementary metrics. For instance, an AEG was classified as CFAE if it complied with both CFE-Mean and CFE-StdDev for NavX classification. Similarly, an AEG was classified as CFAE if it complied with ICL, ACI and SCI for CARTO classification.

2.2. Statistical analysis

All continuous normally distributed variables are expressed as mean \pm standard deviation (SD). Nonparametric paired multiple data were analyzed using the Friedman test with Dunn’s correction. Nonparametric unpaired data were analyzed using the Mann–Whitney test. Categorical data were expressed as percentages and analyzed using the two-sided Yates-corrected Chi-square test. The level of agreement in the CFAE classification performed by the two systems was assessed by the Cohen’s kappa (κ) score [12]. Kappa score within range $0 \leq \kappa < 0.4$ suggests marginal agreement between two metrics; $0.4 \leq \kappa \leq 0.75$ good agreement and; $\kappa > 0.75$ excellent agreement [13]. $P \leq 0.05$ was considered statistically significant.

3. Results

The comparison between CFE-Mean and ICL for each of the 797 AEGs is illustrated in Figure 1, with their respective default thresholds for CFAE detection highlighted. Four quadrants were delimited: two quadrants where ICL and CFE-Mean agreed in terms of categorization, i.e., whether an AEG is fractionated or not fractionated and two quadrants in which they disagreed. Examples of AEGs for each of the quadrants are given.

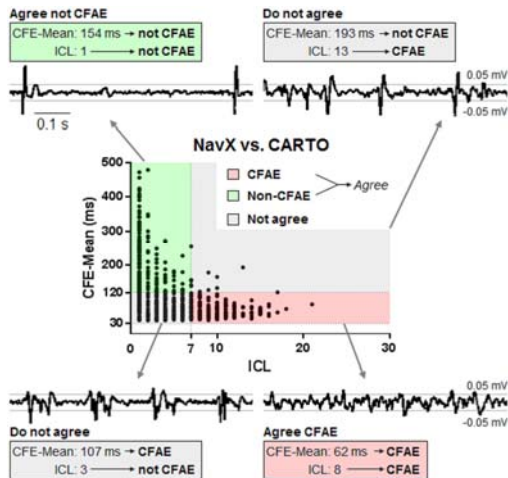


Figure 1. Comparison of CFAE classification between CFE-Mean and ICL measured for all bipolar AEGs.

The CFAE classifications performed by the CFE-Mean ≤ 120 ms suggest that the default threshold for CARTO (ICL ≥ 7) provides high specificity but poor sensitivity for CFAE detection (Table 1, Figure 2A). The optimum threshold found from the ROC curves (ICL ≥ 4) provides optimum sensitivity and specificity for CFAE detection and classification when using NavX as the comparator. The CFAE classifications performed by the ICL ≥ 7 suggest that the default threshold for NavX (CFE-Mean ≤ 120 ms) provides high sensitivity but poor specificity for CFAE detection (Table 1, Figure 2B). Thresholding CFE-Mean ≤ 84 ms provides optimum sensitivity and specificity for CFAE detection and classification when using CARTO as the comparator. The results from the ROC curves suggest that CFE-StdDev ≤ 47 ms, ACI ≤ 82 ms and SCI ≤ 58 ms provide optimum sensitivity and specificity for CFAE detection, when considering the agreement between CFE-Mean and ICL for CFAE classification (Table 2, ROC curves not shown).

Using the default thresholds (NavX: CFE-Mean ≤ 120 ms; CARTO: ICL ≥ 7) NavX classified $69 \pm 5\%$ of the AEGs as CFAEs, while CARTO detected $35 \pm 5\%$

Table 1. Threshold optimization for ICL and CFE-Mean.

Classifier	Thresholds	Sensitivity	1-Specificity	AUROC	P-Value
CFE-Mean ≤ 120 ms	ICL _{Default} ≥ 7	0.492 \pm 0.008	0.050 \pm 0.005	0.852 \pm 0.005	****
	ICL _{Revised} $\geq 3.8 \pm 0.4$	0.777 \pm 0.022	0.162 \pm 0.022		
ICL ≥ 7	CFE-Mean _{Default} ≤ 120 ms	0.958 \pm 0.005	0.552 \pm 0.009	0.755 \pm 0.005	****
	CFE-Mean _{Revised} $\leq 84.1 \pm 0.4$ ms	0.807 \pm 0.010	0.362 \pm 0.006		

AUROC = Area under receiver operating characteristic curve. Values in mean (\pm SD) from the thirty ROC curves according to the CFAE classification by CFE-Mean and ICL. **** P<0.0001.

Table 2. Threshold optimization for CFE-StdDev, ACI and SCI.

Thresholds	Sensitivity	1-Specificity	AUROC	P-value
CFE-StdDev $\leq 46.6 \pm 0.8$ ms	0.905 \pm 0.012	0.185 \pm 0.008	0.877 \pm 0.014	****
ACI $\leq 82.2 \pm 0.3$ ms	0.827 \pm 0.010	0.360 \pm 0.009	0.759 \pm 0.006	****
SCI $\leq 58.6 \pm 0.4$ ms	0.816 \pm 0.012	0.300 \pm 0.009	0.812 \pm 0.005	****

Values in mean (\pm SD) from the thirty ROC curves according to the CFAE classification in agreement by CFE-Mean and ICL. **** P<0.0001.

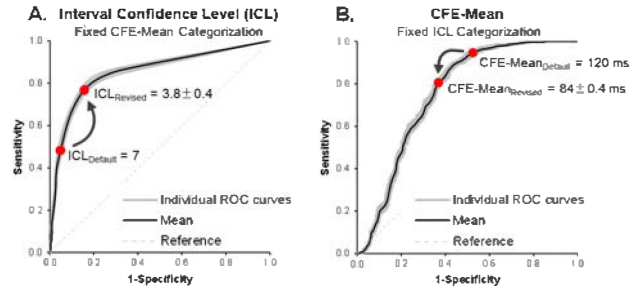


Figure 2. ROC curves and threshold customization for ICL (A) and CFE-Mean (B).

($P < 0.0001$). With the revised thresholds (NavX: CFE-Mean ≤ 84 ms and CFE-StdDev ≤ 47 ms; CARTO: ICL ≥ 4 , ACI ≤ 82 ms and SCI ≤ 58 ms) NavX classified $45 \pm 4\%$, while CARTO detected $42 \pm 5\%$ ($P < 0.0001$). Kappa score between the CFAE categorization performed by NavX and CARTO significantly increased ($P < 0.0001$) from 0.34 ± 0.07 (marginal agreement, $P < 0.0001$) using their default thresholds to 0.45 ± 0.10 (good agreement, $P < 0.0001$) with the revised thresholds.

Figure 3 illustrates the CFAE classification from one patient according to NavX (upper maps) and CARTO (bottom) using their default (left) and revised (right) thresholds. Although the regions marked with CARTO revised thresholds are smaller, they are in agreement with NavX regional classification using the revised thresholds.

4. Discussion and conclusions

The true significance of CFAE in the pathophysiology of AF remains to be determined. NavX and CARTO have been broadly used in automated identification of CFAEs. The results presented here, however, support the perception that CFAE target identification is dependent on the system used during the electrophysiologic procedure [8, 9]. Therefore, each system could identify different atrial regions as ablation targets in the same patient, which could explain the varying outcomes of success rate in CFAE-guided ablation [3-6].

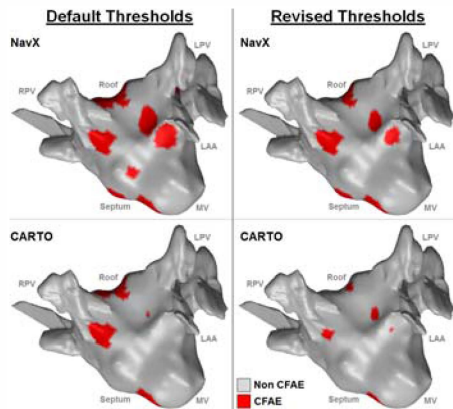


Figure 3. CFAE classification in the LA (anterior view) performed by NavX (upper) and CARTO (bottom) using their default (left) and revised (right) thresholds.

Consequently, discretion is needed when comparing the outcomes in AF ablation incorporating CFAE-targeted approaches in different studies. Minimizing the differences in CFAE classification between NavX and CARTO, therefore, would help to improve the current understanding of the real significance of CFAE in the underlying mechanisms of persAF. We propose customized thresholds – that could be used immediately in different electrophysiological studies using either NavX or CARTO – which optimize the sensitivity and specificity of CFAE detection. These thresholds counterbalance the differences of automated CFAE classification performed by the algorithms embedded in each system and improve the ability of each algorithm to identify CFAEs in agreement with both systems. In other words, a CFAE map created with CARTO utilising the customized CARTO's thresholds will look similar to the one that would have been created with NavX utilising NavX's customized thresholds proposed in this work. This would facilitate direct comparison of persAF CFAE ablation outcome guided by NavX or CARTO in different studies and would allow a common CFAE definition going forward.

The results found in this work are currently being used in an electrophysiologic trial study in our research centre.

Acknowledgements

The authors thank Dr. Guilherme L. Pimentel and Dr. Will Nicolson for their valuable suggestions. This work was supported by the Leicester NIHR Cardiovascular Biomedical Research Unit, UK and Conselho Nacional de Desenvolvimento Científico e Tecnológico (Brazil, Grants N. 200251/2012-0 and 200598/2009-0).

References

[1] Terasawa T, Balk E M, Chung M, Garlitski A C, Alsheikh-Ali A A, Lau J and Ip S 2009 Systematic review: comparative effectiveness of radiofrequency

catheter ablation for atrial fibrillation *Annals of internal medicine* **151** 191-202

- [2] Haissaguerre M, Jais P, Shah D C, Takahashi A, Hocini M, Quiniou G, Garrigue S, Le Mouroux A, Le Metayer P and Clementy J 1998 Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins *N Engl J Med* **339** 659-66
- [3] Nademanee K, McKenzie J, Kosar E, Schwab M, Sunsaneewitayakul B, Vasavakul T, Khunnawat C and Ngarmukos T 2004 A new approach for catheter ablation of atrial fibrillation: Mapping of the electrophysiologic substrate *J Am Coll Cardiol* **43** 2044-53
- [4] Porter M, Spear W, Akar J G, Helms R, Brysiewicz N, Santucci P and Wilber D J 2008 Prospective study of atrial fibrillation termination during ablation guided by automated detection of fractionated electrograms *J Cardiovasc Electrophysiol* **19** 613-20
- [5] Verma A, Lakkireddy D, Wulffhart Z, Pillarisetti J, Farina D, Beardsall M, Whaley B, Giewercer D, Tsang B and Khaykin Y 2011 Relationship Between Complex Fractionated Electrograms (CFE) and Dominant Frequency (DF) Sites and Prospective Assessment of Adding DF-Guided Ablation to Pulmonary Vein Isolation in Persistent Atrial Fibrillation (AF) *Journal of Cardiovascular Electrophysiology* **22** 1309-16
- [6] Verma A, Mantovan R, Macle L, De Martino G, Chen J, Morillo C A, Novak P, Calzolari V, Guerra P G, Nair G, Torrecilla E G and Khaykin Y 2010 Substrate and Trigger Ablation for Reduction of Atrial Fibrillation (STAR AF): a randomized, multicentre, international trial *Eur Heart J* **31** 1344-56
- [7] Monir G and Pollak S J 2008 Consistency of the CFAE phenomena using custom software for automated detection of complex fractionated atrial electrograms (CFAEs) in the left atrium during atrial fibrillation *J Cardiovasc Electrophysiol* **19** 915-9
- [8] Lau D H, Maesen B, Zeemering S, Kuklik P, van Hunnik A, Lankveld T A R, Bidar E, Verheule S, Nijs J, Maessen J, Crijns H, Sanders P and Schotten U 2015 Indices of Bipolar Complex Fractionated Atrial Electrograms Correlate Poorly with Each Other and Atrial Fibrillation Substrate Complexity *Heart Rhythm*
- [9] Almeida T P, Chu G S, Salinet J L, Vanheusden F J, Li X, Tuan J H, Stafford P J, Schilindwein F S and Ng G A Towards the unification of complex fractionated atrial electrogram definition using different electroanatomical mapping systems in persistent atrial fibrillation therapy *Submitted to Medical Engineering & Physics in March 2015*
- [10] Tuan J, Jeilan M, Kundu S, Nicolson W, Chung I, Stafford P J and Ng G A 2011 Regional fractionation and dominant frequency in persistent atrial fibrillation: effects of left atrial ablation and evidence of spatial relationship *Europace* **13** 1550-6
- [11] Fawcett T 2006 An introduction to ROC analysis *Pattern Recognition Letters* **27** 861-74
- [12] Cohen J 1968 Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit *Psychological bulletin* **70** 213-20
- [13] Landis J R and Koch G G 1977 The measurement of observer agreement for categorical data *Biometrics* **33** 159-74