



# Umati: Monitoring Online Dangerous Speech

February and March 2013 report





## Introduction

Hate speech has garnered growing interest in Kenya since the 2007/8 Post Election Violence, in which it seems to have played a role, and because it has been rising again in certain contexts - online for example - in the period leading to our next presidential elections in just over two weeks. In response to this, Umati has conducted a unique, first-ever project to 1) monitor the Kenyan online space for hate speech; 2) analyze the speech for how likely it is to stir violence; 3) find and use non-government ways of countering it.

Under Article 13 of the National Cohesion and Integration Act of 2008, a person who uses speech (including words, programs, images or plays) that is “threatening, abusive or insulting or involves the use of threatening, abusive or insulting words or behaviour commits an offence if such person intends thereby to stir up ethnic hatred, or having regard to all the circumstances, ethnic hatred is likely to be stirred up.” Notably, the Act mentions ethnic hatred only - not hatred based on religion, gender, nationality, sexual preference, or any other group category.

Other Kenyan laws also touch on hate speech, in diverse ways. The 2010 Constitution notes that freedom of expression does not extend to hate speech - but does not define that term. Kenya’s Code of Conduct for political parties (attached to the Political Parties Act) forbids parties to “advocate hatred that constitutes ethnic incitement, vilification of others or incitement to cause harm.”

The law is still imprecise, in other words, and there has been an escalating demand from peacebuilding organisations, politicians, government officials and the general public for more detail on how to define, identify, mitigate, report and deal with hate speech.

This need motivated the Umati project to facilitate easier identification of hate speech, especially the type of hate speech that has a potential to trigger violence so that the violence can be avoided or diminished. The type of hate speech that has the capacity to catalyse violence is known as ‘dangerous speech’.

Dangerous speech does not by itself cause violence, but instead has the capacity to promote or inflame violence - even when people are heavily influenced by speech, they are able to resist its power, and are legally and morally responsible if they commit violence.



Professor Susan Benesch of American University (Washington, DC, USA), an authority on hate speech as a precursor to violence in many countries, defines dangerous speech as speech that has a reasonable possibility of helping to catalyze violence.

She has developed a five-point analytical tool for gauging when violence is likely to be stirred up, to borrow language from the National Cohesion and Integration Act, or as Prof. Benesch puts it, for estimating the dangerousness of a particular speech act in the context in which it was made or disseminated (The impact of speech always depends on the context.) These are factors identified by Professor Benesch that make speech more or less powerful:

- the speaker and his or her influence over an audience (a political, cultural, or religious leader? Someone with a large following of another kind?)
- the audience and its reasons for taking inflammatory speech seriously (already fearful? receiving information mainly from one source?);
- the content in the speech that may be taken as inflammatory (serious offense against what is sacrosanct to another community? Referring to humans as pests or vermin?)

- the social and historical context of the speech (previous clashes between two groups? Competition over land or other resources? hardship?);
- the means of spreading the speech, including the language in which it is expressed (mother tongue?) and medium (a radio station, TV network, or blog that, itself, has influence?).

Note that this list does not include the intent of the speaker. Intent must always be considered when defining a crime or building a case for prosecution, but that is not our purpose. Umati aims above all to prevent violence, and we are also strongly dedicated to freedom of speech. Therefore we seek to prevent dangerous speech and violence by mobilizing civil society, not government regulation or prosecution.

The Umati project seeks to identify and understand the use of dangerous speech in the Kenyan online space, in order to find and use non-government ways to reduce its effects of violence on the ground. To this end, we have created NipeUkweli - an outreach effort to debunk inciteful myths and reduce the possible effects of dangerous speech.



## Goals of Umati

Following the need to define, identify and deal with dangerous speech, the goals of the Umati project are:

1. To correctly define the type of speech that is harmful to the Kenyan society and thus enable all citizens to know the limits and freedoms of their speech. Moreover, it is important to identify dangerous speech early in order to prevent the violence it has the potential to catalyse.
2. To forward calls for help to Uchaguzi ([www.uchaguzi.co.ke](http://www.uchaguzi.co.ke)), a technology-based system that enables citizens to report and keep an eye on election-related events on the ground.
3. To define a process for online hate speech tracking that can be replicated elsewhere.
4. To further civic education on dangerous speech so that Kenyans are more responsible in their communication and interactions with people from differing backgrounds.

It is not the goal of Umati to define the law, or to find and prosecute the perpetrators of dangerous speech. Umati is a civil society project, not a legal or policing body.

### Why Online?

While most projects related to hate speech have been looking at mainstream media, we are aware of the influence—positive and negative—that New Media such as the blogosphere and online forums had during the 2007 Post Election Violence in Kenya. Therefore, our flagship project seeks to monitor and report, for the first time, the role New Media plays on a Kenyan election.



## Monitoring Process

Beginning in September 2012, the Umati project has been monitoring online content and recorded incidences of hate and dangerous speech - categorising them according to dangerousness, using the five criteria noted above. We plan to continue monitoring through May 2013, with a final analysis report to be released by June 2013.

This process is being carried out by five monitors, representing the four largest ethnic groups in Kenya, with the fifth monitor focusing on Kiswahili, the national language and Sheng, which is a slang dialect mixing Kiswahili and English.

Cited incidences of hate speech are translated from vernacular to the country's official language, English. The monitors check blogs written in their vernacular language, blogs in English, Facebook pages and groups, Twitter timelines, online newspapers and video streams of the major media houses in Kenya.

As of February 2013, we added a sixth monitor from the Somali community.

An additional team of five monitors were also added to cover the weekends.



How to identify dangerous speech

Picture source: <http://assets.vice.com/content-images/contentimage/no-slug/435640277111c164291446a57>



## How Do You Identify Dangerous Speech?

### 1. It is targeted at a group of people and not a single person.

Dangerous speech is harmful speech that calls the audience to condone or take part in violent acts against a group of people.

From our research we have seen that dangerous speech in the Kenyan on-line space occurs along various lines including tribal, religious, gender, political and racial lines.

It's important to note that an ugly or critical comment about an individual - a politician, for example - is not hate speech unless it targets that person as a member of a group.

Hate speech is directed at a group, or at a person as part of a group: a tribe, religion, women, homosexuals etc.

During election periods, it is not uncommon for negative statements to be made against politicians and other influential personalities. This is a normal part of the political process, as long as the statements do not constitute defamation, threats, hate speech, or dangerous speech.

### 2. May contain one of the hallmarks/pillars of dangerous speech

Three hallmarks common in several dangerous speech statements are :

- Compares a group of people with animals, insects or vermin
- Suggests that the audience faces a serious threat or violence from another group ("accusation in a mirror")
- Suggests that some people from another group are spoiling the purity or integrity of the speakers' group

Note that a speech statement can still be dangerous despite not having any of these three mentioned pillars of dangerous speech. The hallmarks serve as a diagnostic tool to identify some dangerous speech, since they are commonly (but not universally) found in it.

Also note the converse: a hallmark does not automatically make speech dangerous. As an example, if a mother tells her daughter to stop seeing a boy from another community, and calls the boy by the name of an animal, the speech is almost certainly not dangerous since the daughter will not react with violence against the boy or his community.

### 3. Contains a call to action

Dangerous speech often encourages the audience to condone or commit violent acts on the targeted group. The six calls to action common in dangerous speech are, calls to:

- discriminate,
- loot,
- riot,
- beat,
- forcefully evict, and
- kill.

Statements with a call to beat, forcefully evict or to kill a group amount to an extremely dangerous speech statement, the most vitriolic category of hate speech statement.



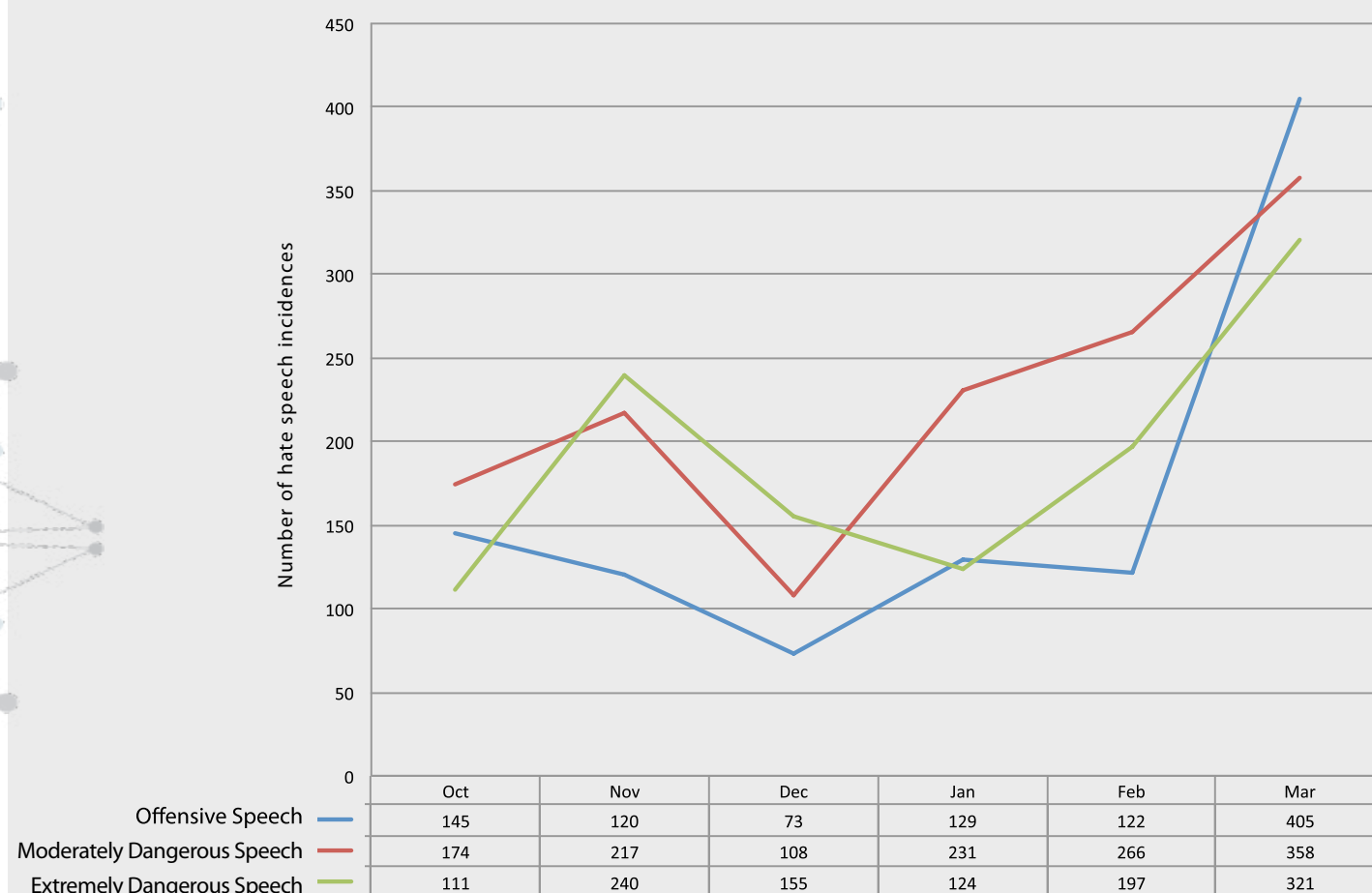
Notable trends in February & March

Picture source: <http://cafod.files.wordpress.com/2013/03/polling-clerks-kenya-elections.jpg>



## Notable Trends

### 1. Increased incidence of hate and dangerous speech



Between November 2012 and January 2013, Umati saw a declining number of extremely dangerous speech statements (category 3 statements) on the Kenyan online space.

However, in the weeks leading up and following the elections, we see an increase in dangerous speech.

Looking at the hate speech categories across the months (Chart B), there is a clear rise in March, the month of the General Elections.

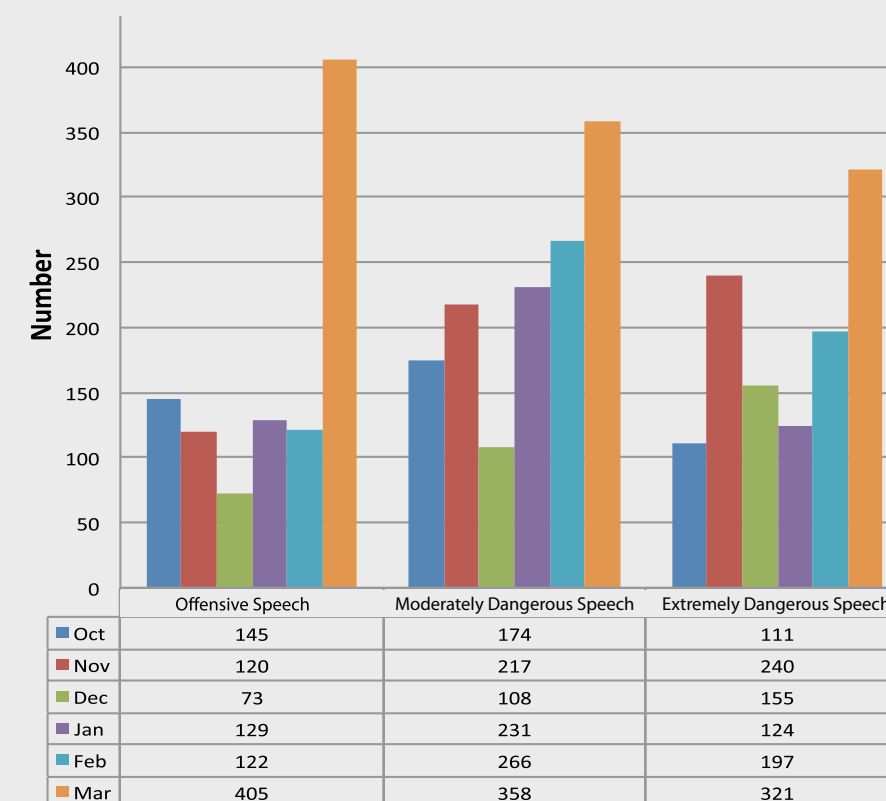


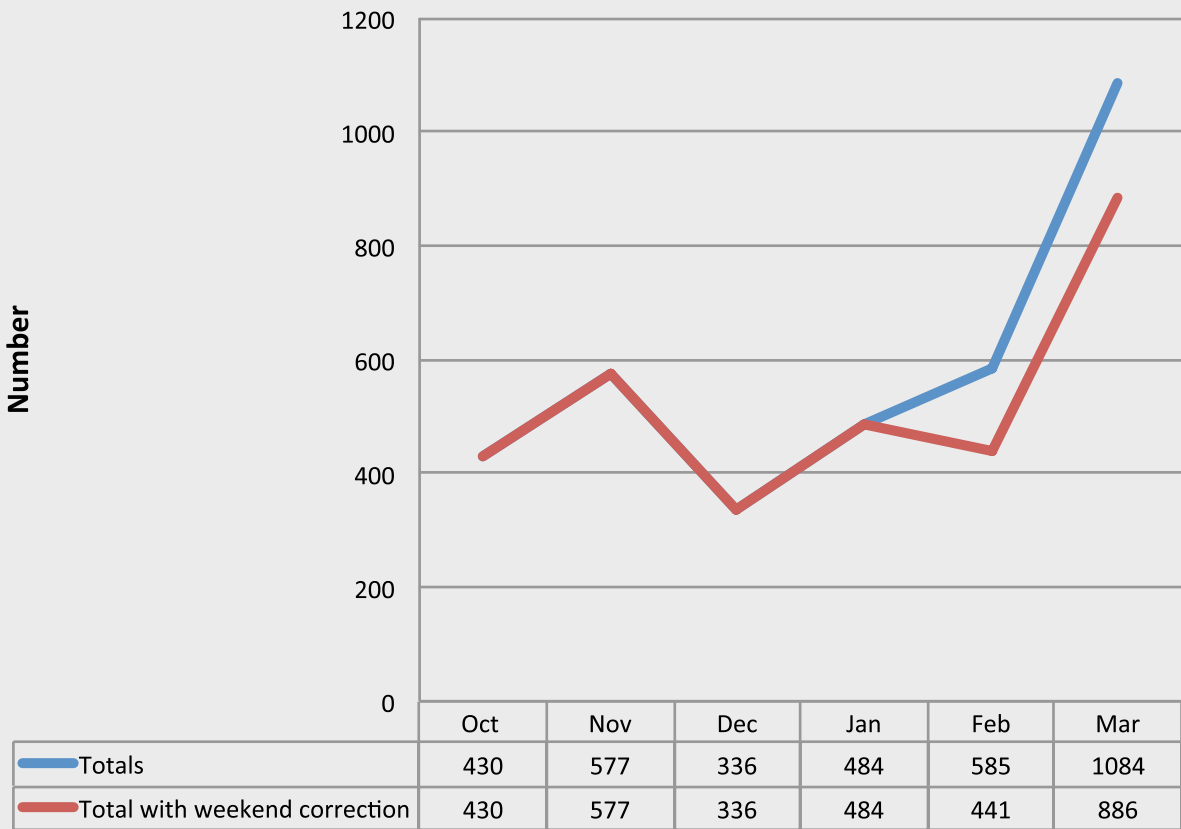
Chart B

Reasons for the increase can be credited to the following:

- March 4 presidential election, tallying, court proceedings and final decision of the Supreme Court. The 2013 elections were largely peaceful, however much of the "violence" shifted to the online space, especially on Facebook and Twitter.
- Expansion of Umati project: In February, a team of five monitors was added to monitor on Saturdays and Sundays. This increased the volume of hate speech we collected.

1b. Increased incidences of hate and dangerous speech

Graph comparing total with weekend correction



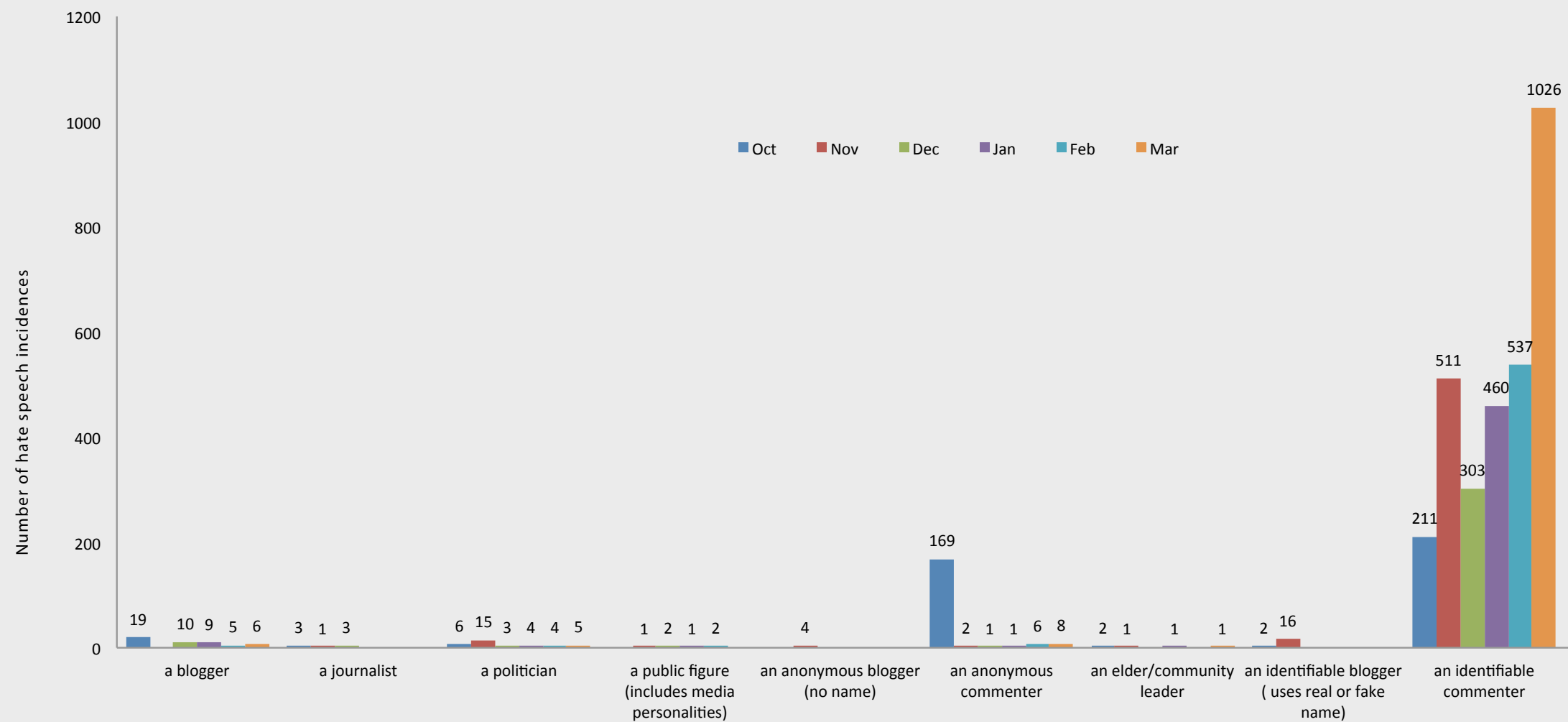
If we adjust the totals to cater for the increased volume due to the weekend monitoring, we still see a rise in hate and dangerous speech statements in March 2013.

Even after discarding weekend monitoring, there is a doubling of statements collected in March, as compared to those collected in February.

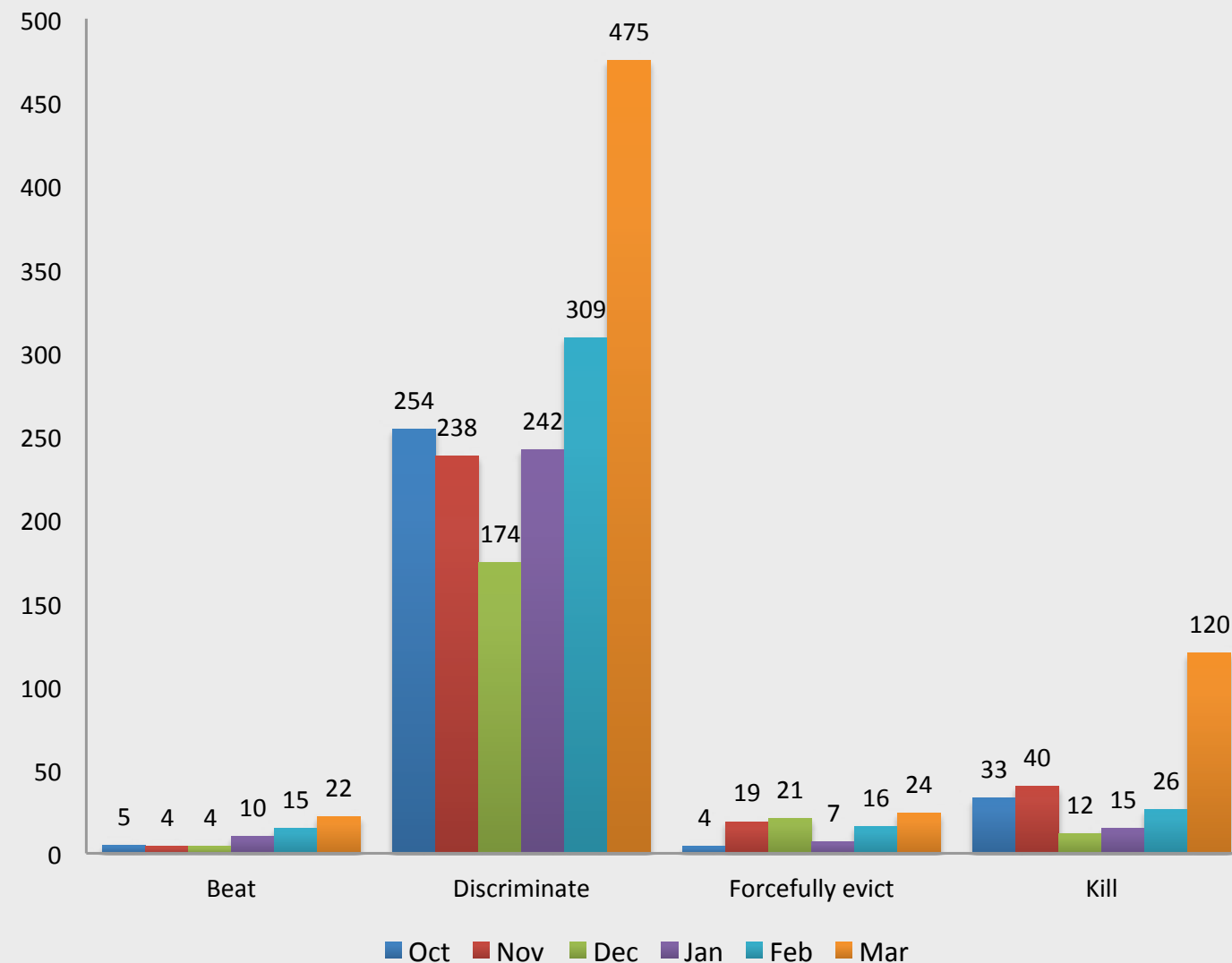
The totals for data collected over the February and March weekends were:

February	144
March	198

## 2. Identifiable commenters remain the most active users of dangerous speech



### 3. Sharp increase in calls to kill



Calls to action are used to determine whether a statement is an extremely dangerous statement. The three calls to action to determine this are:

- call to kill
- call to forcefully evict
- call to beat

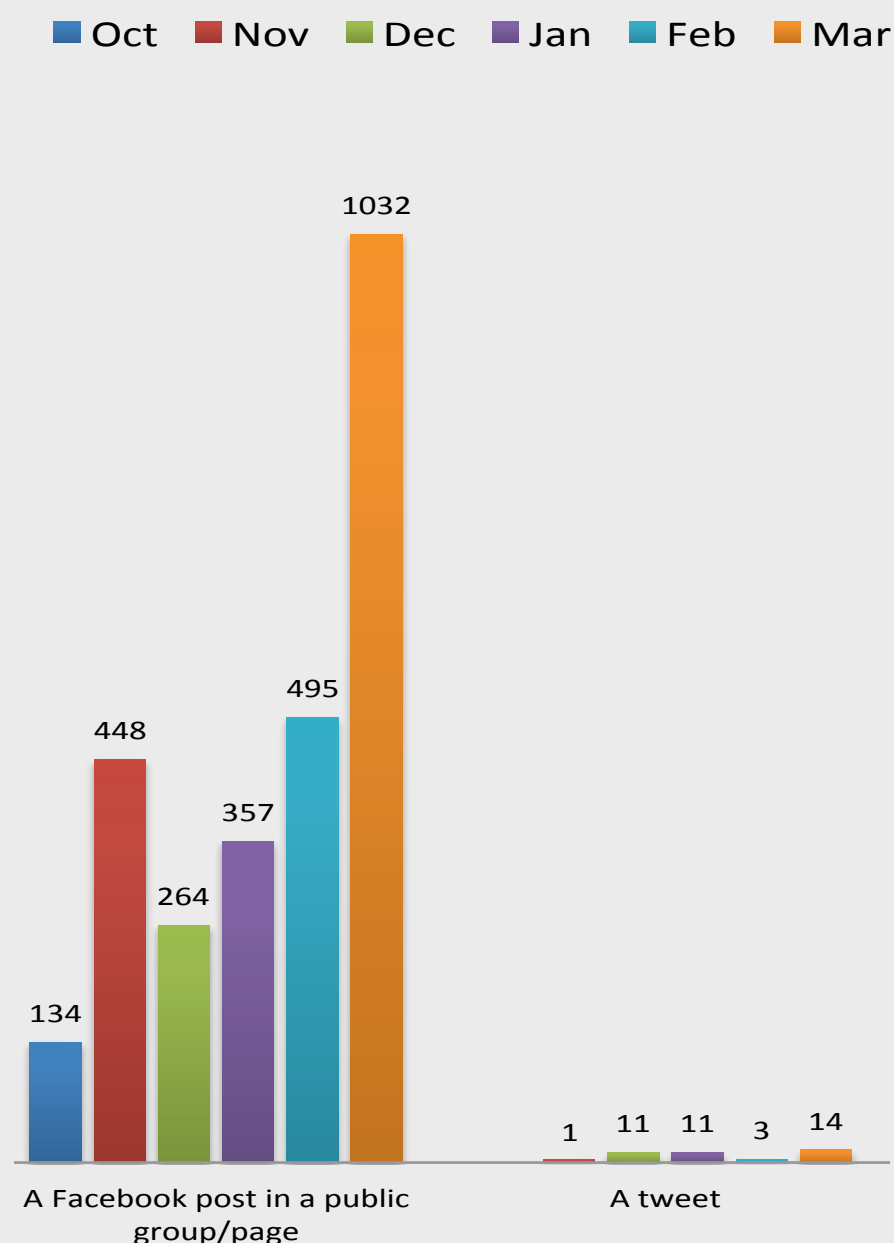
Statements encouraging the audience to beat, forcefully evict or kill a group of people, or condone such behaviour are extremely dangerous speech statements.

In previous months, the most rampant call to action has been the call to discriminate, whether via insults or stereotypes etc. This trend is maintained in February and March.

Of note, is the sharp increase in calls to kill that has been witnessed in March.



#### 4. Non-increase in Twitter activity



During the pre and post election period, Kenyans on Twitter have been seen to be very vocal in preserving the country's image. Foreign journalists believed to have written exaggerated or untrue articles about the situation in the country, have received massive online attacks from the Kenyan Twitter community. One popular cyber attack led to the resurgence of the #sometellTellCNN hashtag, which targeted at a journalist who insinuated that there will be war after the March 4th election. Other hash tags against foreigners seen to taint Kenya's image include #someoneTellBBC, #someoneTellFrance24, #someoneTellBotswana, #someoneTellNigeria.

These cyber attacks implied that Twitter would see a rise in hate and dangerous speech conversations. Data from Umati however showed no significant increase. It is apparent that Facebook remains the platform where most hate speech conversations occur, both in public and private groups.

There has however been concern over to what extent Twitter and Facebook onslaughts against foreign journalists on can amount to dangerous speech, as indeed foreign journalists can be a victimised group of people, that same way a particular tribe can be.

Though speech that attacks a particular person does not amount to "Dangerous Speech" since it is not likely to catalyze mass violence, incitement to violence against individuals can also be dangerous, and some foreign journalists have indeed received death threats. It is also important to note that other types of speech exist that are punishable by law: defamation, threats and cyber bullying.



For more information on this project, contact



Umati Project Team

iHub Research

Nairobi, Kenya

[umati@ihub.co.ke](mailto:umati@ihub.co.ke)

[www.research.ihub.co.ke](http://www.research.ihub.co.ke) | Twitter: [@iHubResearch](https://twitter.com/iHubResearch)

