Capítulo 4

# 4 Niveles de Confianza

En los sistemas de reconocimiento automático de habla (sistemas ASR), los niveles de confianza determinan cuando una frase debe ser aceptada, rechazada, o confirmada. En este capítulo se describen los niveles de confianza del reconocedor SpeechWorks y se propone un método para definir los niveles de confianza adecuados para un sistema específico. Por último, se hace una comparación de desempeño de los experimentos base y el desempeño del sistema con los niveles de confianza generados por el método propuesto en este capitulo.

## 4.1 Introducción

Debido a que los sistemas ASR aún carecen de perfección, los niveles de confianza son de gran utilidad en el proceso de aceptación o rechazo de frases. A través de una probabilidad que representa el nivel de confianza, se decide si una palabra o frase es aceptada, rechazada o confirmada. En el diseño e implementación de sistemas ASR, el proceso de niveles de confianza es muy útil. Considere un ejemplo de un sistema que ofrece el servicio de conmutador automático, y que hace la siguiente pregunta: "Por favor *diga el nombre de la persona con la que desea hablar*", y el usuario da como respuesta "*Si, me gustaría hablar con Alcira Vargas, por favor*". Para efectos de este ejemplo, el reconocedor reporta un nivel de confianza de 85% en que la respuesta es "Alcira Vargas". Con el nivel de confianza obtenido se pueden tomar dos decisiones: Aceptar la frase a pesar de la falta de seguridad, o confirmar la frase para asegurar el reconocimiento, lo cual corresponde a los desarrolladores de aplicaciones decidir que hacer con esta probabilidad.

## 4.2 Niveles de Confianza en SpeechWorks

El reconocedor de SpeechWorks utiliza dos umbrales para aceptar, rechazar, o confirmar frases: el umbral de alta confianza y el umbral de baja confianza. Los niveles de confianza en SpeechWorks tienen un rango de 0 a 1000; entre más alto sea el nivel, más alta será la confianza del reconocedor. En la figura 4-1 se ilustran los dos umbrales de confianza del reconocedor SpeechWorks.

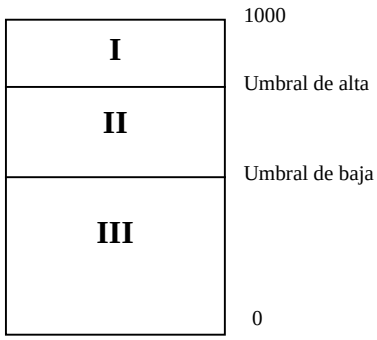



Figura 4-1 Niveles de Confianza en SpeechWorks.

Si el nivel de confianza de la frase del locutor se encuentra en la región I, la frase es aceptada, mientras que si el nivel de confianza se encuentra en la región II, la frase es confirmada. Las frases con niveles de confianza en la región III son siempre rechazadas. Si el umbral alto es igual al umbral bajo, no hay confirmación dentro de la aplicación. En la figura 4-2, se muestra un diagrama de flujo para los niveles de confianza en SpeechWorks.


Figura 4-2 Diagrama de decisión de los niveles de confianza en SpeechWorks.

### 4.2.1 Discusión

De acuerdo a la figura 4-1, cuando se incrementa el umbral de alta confianza, el porcentaje de confirmación incrementa, pero el porcentaje de aceptación disminuye. Por otro lado, cuando se incrementa el umbral de baja confianza, el porcentaje de confirmación disminuye, pero el porcentaje de rechazo de las medidas dentro y fuera del vocabulario aumenta. Debido a esta situación, la definición de umbrales en un sistema ASR tiene una repercusión considerable en su desempeño. Además, los umbrales de confianza varían dependiendo de la aplicación. Por lo tanto, queda claro que se debe desarrollar un mecanismo que genere los niveles de confianza más apropiados por los requerimientos de una aplicación específica. La siguiente sección explica un método para realizar esta tarea.

## 4.3 Optimización de Umbrales de Confianza

Debido a que la definición de niveles de confianza esta en función de una aplicación específica, en esta sección se propone un método que genera umbrales de confianza apropiados a partir de la figura de mérito (FOM) explicada en la sección 3.3.1. Considerando el hecho de que los niveles de confianza del reconocedor SpeechWorks tienen valores en el rango de 0 a 1000, el cálculo de optimización de umbrales de confianza está definido por la siguiente ecuación:

$$NC_{optimos} = \max_{j>i} FOM(i, j) \qquad (4.1)$$

$$para \quad i = 0,10,20,...,100$$

$$j = 0,10,20,...,100$$

En la ecuación 4.1, *i* representa el umbral de baja confianza y *j* representa el umbral de alta confianza. La salida es una gráfica en 3D, donde los ejes X y Z representan los umbrales de alta y baja confianza respectivamente, y el eje Y representa la figura de mérito (FOM).

A continuación, se hace una evaluación de los experimentos base, usando la metodología de evaluación propuesta en la sección 3.3. También, se calculan los umbrales de confianza óptimos para tales experimentos.

## 4.4 Evaluación de los experimentos base

Todas las evaluaciones realizadas en esta tesis son fuera de tiempo real. En esta sección se mide el desempeño de dos experimentos en el corpus de la estructura experimental. Los experimentos consisten en identificar una palabra clave en las frases. El primero usa solo vocabulario (no permite palabras no clave) y el segundo usa una gramática predefinida (gramática especial de CONMAT que permite palabras no clave). El esquema que describe el experimento de solo vocabulario se ilustra en la figura 4-3, y la gramática que describe el segundo experimento se ilustra en la figura 4-4, donde *N* el número de las palabras clave. Algunos ejemplos que permite el experimento de gramática predefinida se muestran en la tabla 4-1. La gramática de CONMAT en su forma BNF se describe en el apéndice A1.

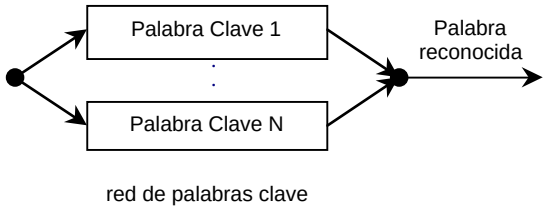



Figura 4-3 Esquema del experimento de solo vocabulario.

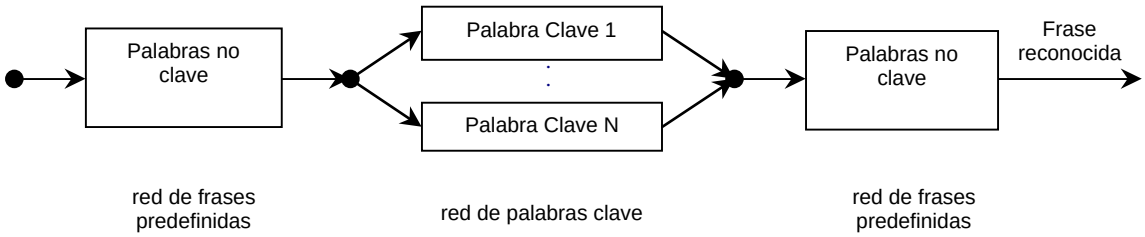



Figura 4-4 Esquema del experimento que usa una gramática predefinida para permitir frases combinando palabras clave con palabras no clave.

| FRASE |
|---|
| Por favor comuníqueme con la biblioteca |
| Quiero hablar a sorteos |
| Me comunica con el doctor Jorge Welti por favor |
| Quiero que me comunique al departamento de electrónica |

Tabla 4-1 Ejemplos de frases permitidas en la gramática predefinida de CONMAT

Los resultados de los experimentos ilustrados en las figuras 4-3 y 4-4, con base a desempeño en el reconocimiento se muestran en la tabla 4-2 y los resultados con base a costo computacional se muestran en la tabla 4-3. Los niveles de confianza usados en esta evaluación son los que usa CONMAT actualmente (550 para el umbral bajo y 900 para el umbral alto). Estos niveles de confianza fueron definidos subjetivamente a través de la experiencia en la aplicación.

| EXPERIMENTO | rr_in | ca_in | cc_in | fa_in | fr_in | fc_in | crout | faout | fcout | FOM |
|---|---|---|---|---|---|---|---|---|---|---|
| Solo vocabulario | 0.811 | 0.581 | 0.214 | 0.018 | 0.087 | 0.100 | 0.385 | 0.115 | 0.500 | 91.03% |
| Gramática predeterminada | 0.857 | 0.579 | 0.258 | 0.009 | 0.067 | 0.087 | 0.365 | 0.115 | 0.519 | 91.59% |

Tabla 4-2 Desempeño de los experimentos base.

| EXPERIMENTO | COSTO POR SEGUNDO DE HABLA |
|---|---|
| Solo vocabulario | 0.97 |
| Gramática predefinida | 1.46 |

Tabla 4-3 Costo computacional de los experimentos base.

## 4.4.1 Análisis de Resultados

En la tabla 4-2 se puede observar que el FOM de los experimentos base presenta una diferencia mínima. Esto se debe principalmente a que el 78% de las frases del corpus contienen palabras clave y a que la gramática perjudica el porcentaje de rechazos correctos (*crout*), debido a que siempre intentar reconocer una palabra clave. Para hacer un análisis comparativo de los experimentos base, se hace uso de los tipos de habla clasificados en nodos terminales (ver sección 3.2). Los resultados se muestran en las tablas 4-4 y 4-5 con una descripción de los nodos terminales en la tabla 4-6.

| NODO TERMINAL | Rr_in | ca_in | cc_in | fa_in | fr_in | Fc_in | crout | faout | fcout | # FRASES |
|---|---|---|---|---|---|---|---|---|---|---|
| T1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | NaN* | NaN | NaN | 3 |
| T2 | 0.108 | 0.027 | 0.054 | 0.108 | 0.405 | 0.405 | NaN | NaN | NaN | 37 |
| T3 | NaN | NaN | NaN | NaN | NaN | NaN | 0.400 | 0.169 | 0.431 | 65 |
| T4 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000 | 0.000 | 1.000 | 6 |
| T5 | NaN | NaN | NaN | NaN | NaN | NaN | 0.424 | 0.030 | 0.545 | 33 |
| T6 | 0.900 | 0.500 | 0.400 | 0.100 | 0.000 | 0.000 | NaN | NaN | NaN | 10 |
| T7 | 0.880 | 0.639 | 0.226 | 0.008 | 0.060 | 0.068 | NaN | NaN | NaN | 399 |

Tabla 4-4 Resultados de reconocimiento del experimento base con solo vocabulario.

A partir de los resultados de reconocimiento del experimento con solo vocabulario, se pueden hacer las siguientes afirmaciones:

- Tiene buen desempeño cuando existen únicamente palabras clave (T6 y T7).
- Es deficiente para rechazar frases con solo palabras no clave (T3, T4 y T5).
- Tiene muy bajo desempeño con frases que contienen palabras no clave combinadas con palabras clave (T1 y T2).

| NODO TERMINAL | rr_in | ca_in | cc_in | fa_in | fr_in | Fc_in | crout | faout | fcout | # FRASES |
|---|---|---|---|---|---|---|---|---|---|---|
| T1 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | NaN | NaN | NaN | 2 |
| T2 | 0.455 | 0.091 | 0.364 | 0.000 | 0.182 | 0.364 | NaN | NaN | NaN | 11 |
| T3 | NaN | NaN | NaN | NaN | NaN | NaN | 0.400 | 0.169 | 0.431 | 65 |
| T4 | NaN | NaN | NaN | NaN | NaN | NaN | 0.000 | 0.000 | 1.000 | 6 |
| T5 | NaN | NaN | NaN | NaN | NaN | NaN | 0.364 | 0.030 | 0.606 | 33 |
| T6 | 1.000 | 0.500 | 0.500 | 0.000 | 0.000 | 0.000 | NaN | NaN | NaN | 10 |
| T7 | 0.877 | 0.594 | 0.261 | 0.010 | 0.063 | 0.073 | NaN | NaN | NaN | 399 |
| T8 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.000 | NaN | NaN | NaN | 1 |
| T9 | 0.769 | 0.654 | 0.115 | 0.000 | 0.115 | 0.115 | NaN | NaN | NaN | 26 |

Tabla 4-5 Resultados de reconocimiento del experimento base con gramática predefinida.

NaN: No existe un número debido a que el numero de frases para este caso es cero, por lo que 0/0 es definido como Nan.

| NODO TERMINAL | DESCRIPCION |
|---|---|
| T1 | No analizable + palabras fuera + palabras clave + diacríticos |
| T2 | No analizable + palabras fuera + palabras clave |
| T3 | No analizable + diacríticos |
| T4 | No analizable + palabras fuera + diacríticos |
| T5 | No analizable + palabras fuera |
| T6 | Analizable + palabra clave + diacríticos |
| T7 | Analizable + palabra clave |
| T8 | Analizable + lenguaje natural + diacríticos |
| T9 | Analizable + lenguaje natural |

Tabla 4-6 Descripción de los nodos terminales para aplicaciones con gramática.

A partir de los resultados de reconocimiento del experimento de gramática predefinida, se pueden hacer las siguientes afirmaciones:

- Tiene buen desempeño cuando existen únicamente palabras clave (T6 y T7).
- Tiene desempeño aceptable con frases que contienen palabras no clave, pero que son analizables por el parser del reconocedor (T8 y T9).
- Tiene bajo desempeño para rechazar frases con solo palabras no clave (T3, T4 y T5).
- Tiene bajo desempeño cuando se enfrenta con frases que contienen palabras no clave y que no son analizables por el parser reconocedor (T1 y T2).

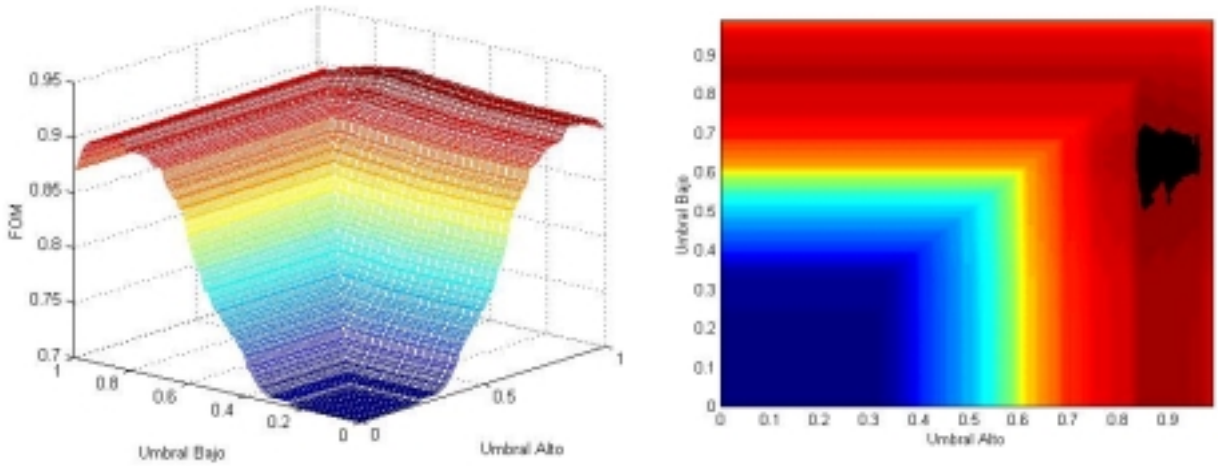
El experimento con gramática predefinida es un buen punto de partida para hacer un estudio sobre habla fuera del vocabulario. Por lo tanto, hay dos situaciones principales que hay que abordar en los siguientes capítulos: el rechazo correcto de frases con solo palabras no clave (T2, T3 y T4) y precisión en el reconocimiento de frases que contienen palabras no clave combinadas con palabras clave (T1, T2, T8 y T9), sin disminuir el desempeño de las frases con solo la palabra clave (T6 y T7).

## 4.5 Optimización de Umbrales de Confianza en los Experimentos Base

Usando el calculo de niveles de confianza explicado en la sección 4.3, los umbrales de confianza óptimos para el experimento que usa solo vocabulario fueron **620** y **860**. Los niveles de confianza óptimos para el experimento que usa una gramática predefinida fueron **650** y **850**. Los resultados de reconocimiento se muestran en la tabla 4-7 y en la figura 4-5.

| EXPERIMENTO | rr_in | ca_in | cc_in | fa_in | fr_in | fc_in | crout | faout | fcout | FOM |
|---|---|---|---|---|---|---|---|---|---|---|
| Solo vocabulario | 0.811 | 0.581 | 0.214 | 0.018 | 0.087 | 0.100 | 0.385 | 0.115 | 0.500 | 91.03% |
| N-Solo vocabulario | 0.811 | 0.619 | 0.149 | 0.018 | 0.143 | 0.071 | 0.606 | 0.115 | 0.279 | 91.47% |
| Gramática predeterminada | 0.857 | 0.579 | 0.258 | 0.009 | 0.067 | 0.087 | 0.365 | 0.115 | 0.519 | 91.59% |
| N-Gramática predeterminada | 0.857 | 0.630 | 0.167 | 0.011 | 0.149 | 0.042 | 0.635 | 0.125 | 0.240 | 91.88% |

Tabla 4-7 Resultados de los niveles de confianza optimizados (N-) comparados con los resultados con los niveles de confianza de los experimentos base.




(a) (b)

Figura 4-5 Umbrales de confianza apropiados para el experimento de gramática predefinida, la figura (a) es una vista desde un costado y (b) es una vista desde la parte superior, la mancha de color negro representa la mejor región para definir niveles de confianza.

## 4.6 Resumen

En este capítulo se explicó la utilidad de los niveles de confianza. Se explicaron los niveles de confianza del reconocedor SpeechWorks y se propuso un método para optimizar los umbrales de confianza basado en las medidas de desempeño del reconocedor. Se destaca que los umbrales de confianza pueden variar de acuerdo al contexto de la aplicación. Con base en una metodología de evaluación definida en el capítulo anterior, se midió el desempeño de dos experimentos que son considerados como punto de partida (solo vocabulario y gramática predefinida de CONMAT). Se concluye que para que un sistema de identificación de palabras clave tenga un buen desempeño, debe considerar un estudio más a fondo de las frases que contienen palabras no clave combinadas con palabras clave. También, debe considerar el rechazo correcto de frases que contienen solo palabras no clave.

En el siguiente capítulo se investiga y experimenta la técnica de identificación de palabras clave usando fonemas como fillers, en un esfuerzo por mejorar el desempeño cuando las frases contienen palabras no clave.