

Información Mutua entre series temporales

Marcos Martínez Jiménez y Miguel Carreras Lahoz

November 5, 2021

1 Preprocesamiento de series de datos

Antes de analizar las series de datos se realizaron las siguientes etapas de preprocesamiento:

1. Normalización
2. Binarización (o SAX / patrones de orden)
3. Enventanado
4. Codificación en palabras

1.1 Normalización

Las series de datos consideradas abarcan distintos rangos de voltaje (Figura 1) por lo que se realizó una z-normalización para que sus rangos fueran más consistentes y se pudiera aplicar un tratamiento único en la medida de lo posible.

Para una serie de datos $V(t)$, la z-normalización consiste en generar la serie $V_z(t)$:

$$V_z(t) = \frac{V(t) - \mu_{V(t)}}{\sigma_{V(t)}} \quad (1.1)$$

Así, la media y desviación estándar de $V_z(t)$ quedan como:

$$\begin{aligned} \mu_{V(t)} &= 0 \\ \sigma_{V(t)} &= 0 \end{aligned} \quad (1.2)$$

La Figura 1 muestra un intervalo de las series de datos normalizadas. Se puede apreciar que el rango de voltajes ahora es mucho más consistente.

1.2 Binarización

La binarización de las series de datos se realizó mediante un umbral de voltaje u_V , de forma que la serie binarizada $V_b(t)$ viene dada por:

$$V_b(t) = \begin{cases} 1, & V_z(t) \geq u_V \\ 0, & V_z(t) < u_V \end{cases} \quad (1.3)$$

La elección de u_V se realizó en base a las distribuciones de $V_z(t)$ (Figura 2) y mediante una comprobación visual de que el valor obtenido era capaz de discriminar los spikes de la señal de

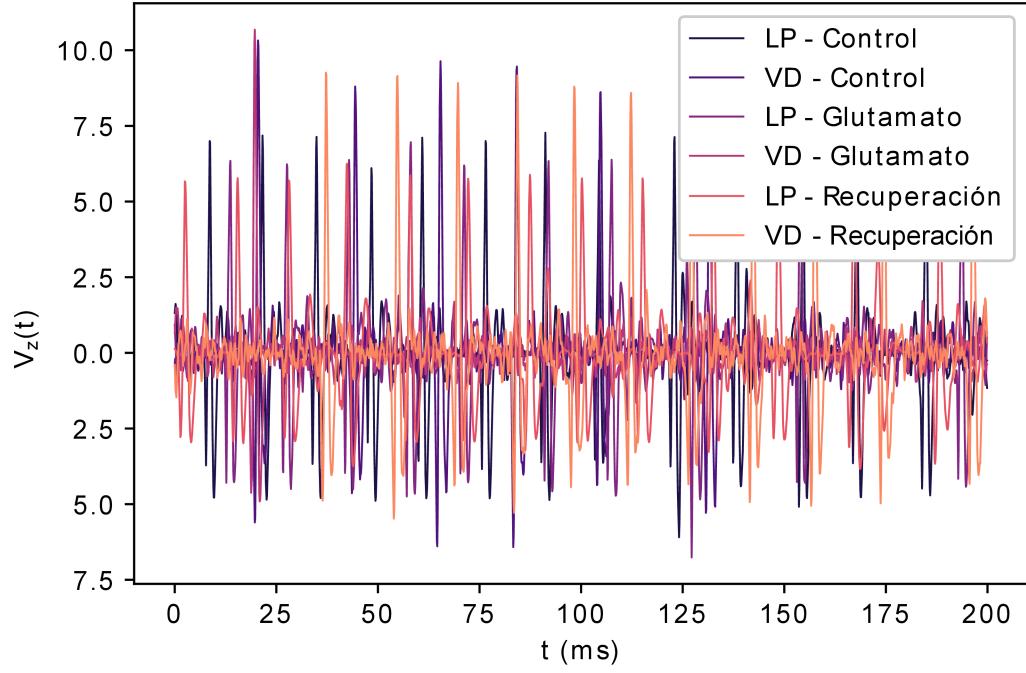
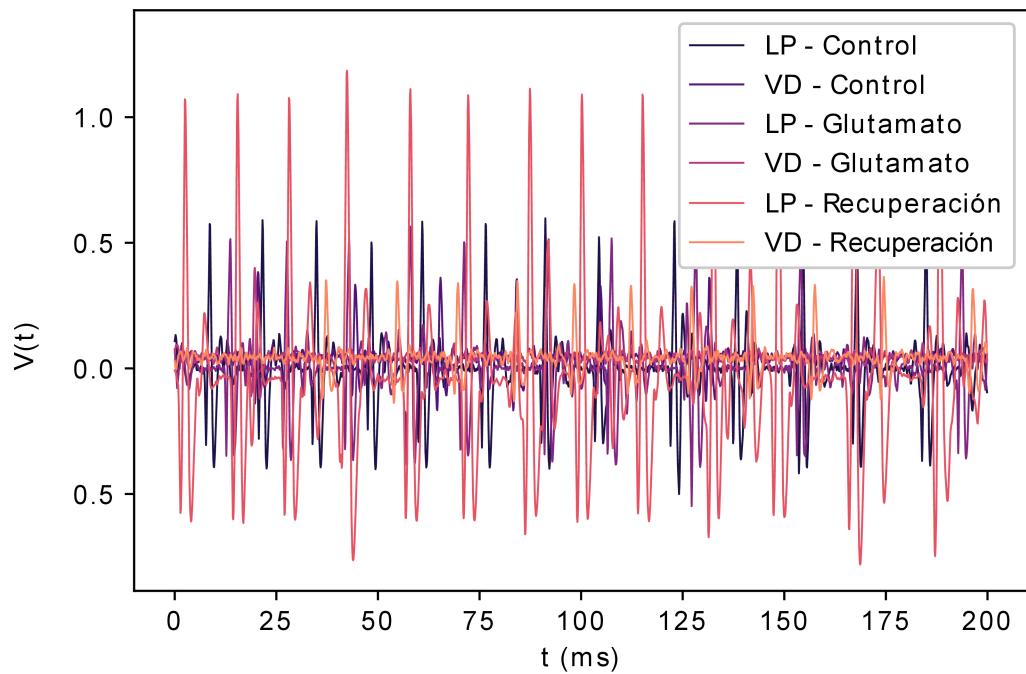


Figura 1: Series de datos sin normalizar y normalizadas. Para cada serie se muestra un intervalo diferente de 200ms en el que aparecen disparos neuronales. El rango de voltaje es mucho más variable en las series sin normalizar.

fondo (Figura 3). La Figura 2 muestra que todas las series tienen una distribución multimodal con un pico entorno a $V_z = 0$, correspondiente a la señal de fondo, otro entorno a $V_z > 4$, correspondiente a los picos de los disparos neuronales, y otro entorno a $V_z < -2$ que corresponde a los valles. La inspección visual de las series de datos confirmó que $u_V = 4$ era un umbral adecuado para detectar los disparos (Figura 3).

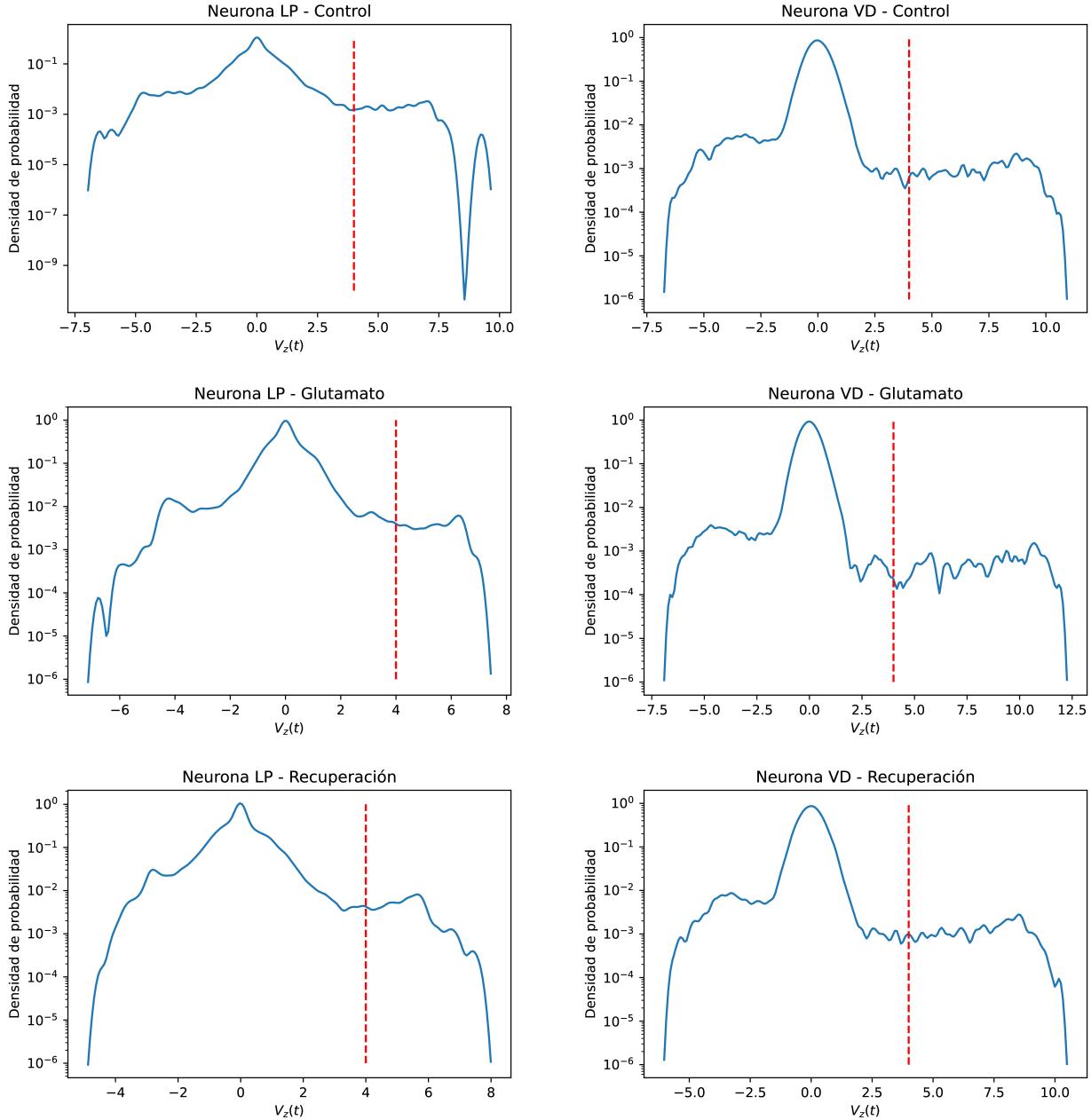


Figura 2: Distribución de voltaje de las series de datos estimada mediante métodos de densidad del núcleo. Todas las series presentan distribuciones multimodales con un punto de separación entorno a $V_z = 4$. La densidad de probabilidad estimada se muestra en escala logarítmica para que la distribución multimodal sea apreciable.

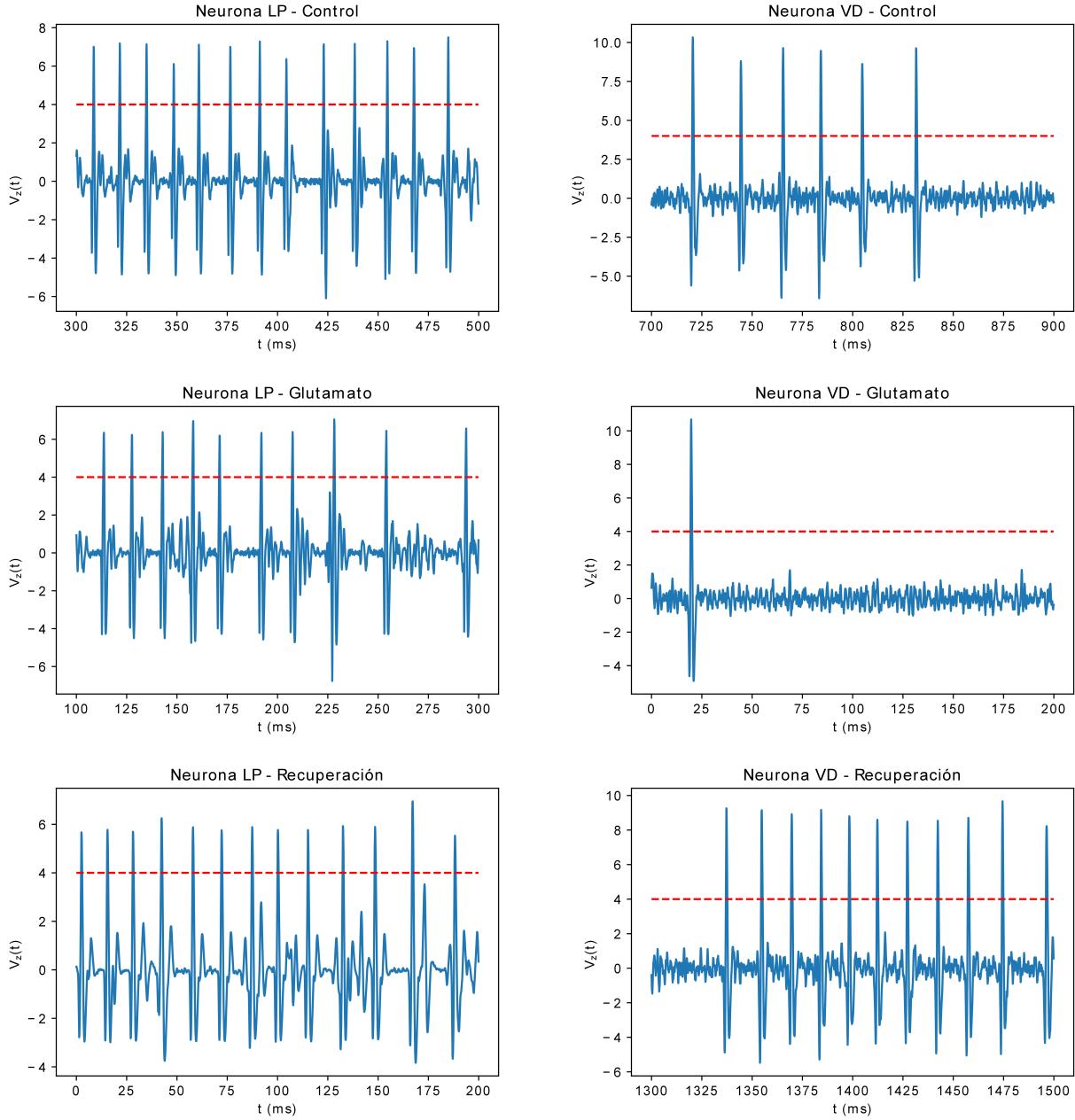


Figura 3: Series de datos de las neuronas LP y VD en control, inyección de glutamato y recuperación. Se muestra el umbral $u_V = 4$, confirmando que detecta adecuadamente los disparos neuronales.

1.3 Enventanado

Una vez binarizadas, las series $V_b(t)$ se enventanaron dando lugar a $V_b^*(w)$ mediante el siguiente esquema. Sea Δt el tamaño de ventana y $[t_i, t_i + \Delta t]$ el intervalo temporal correspondiente a la ventana w_i :

$$V_b^*(w_i) = \begin{cases} 1, & 1 \in \{V_b(t^*) : t^* \in [t_i, t_i + \Delta t]\} \\ 0, & 1 \notin \{V_b(t^*) : t^* \in [t_i, t_i + \Delta t]\} \end{cases} \quad (1.4)$$

Para decidir el rango de valores Δt a utilizar se analizó el número de disparos múltiples incluidos en una sola ventana (Figura 4) y el número de disparos que quedan repartidos entre distintas ventanas (Figura 5) en función de Δt . La Figura 4 muestra que $\Delta t \leq 15$ da lugar a una frecuencia razonablemente baja de disparos múltiples. A su vez, la Figura 5 muestra que la frecuencia de disparos repetidos es bastante baja independientemente de Δt por lo que se escogió un intervalo de Δt de $[1ms, 15ms]$.

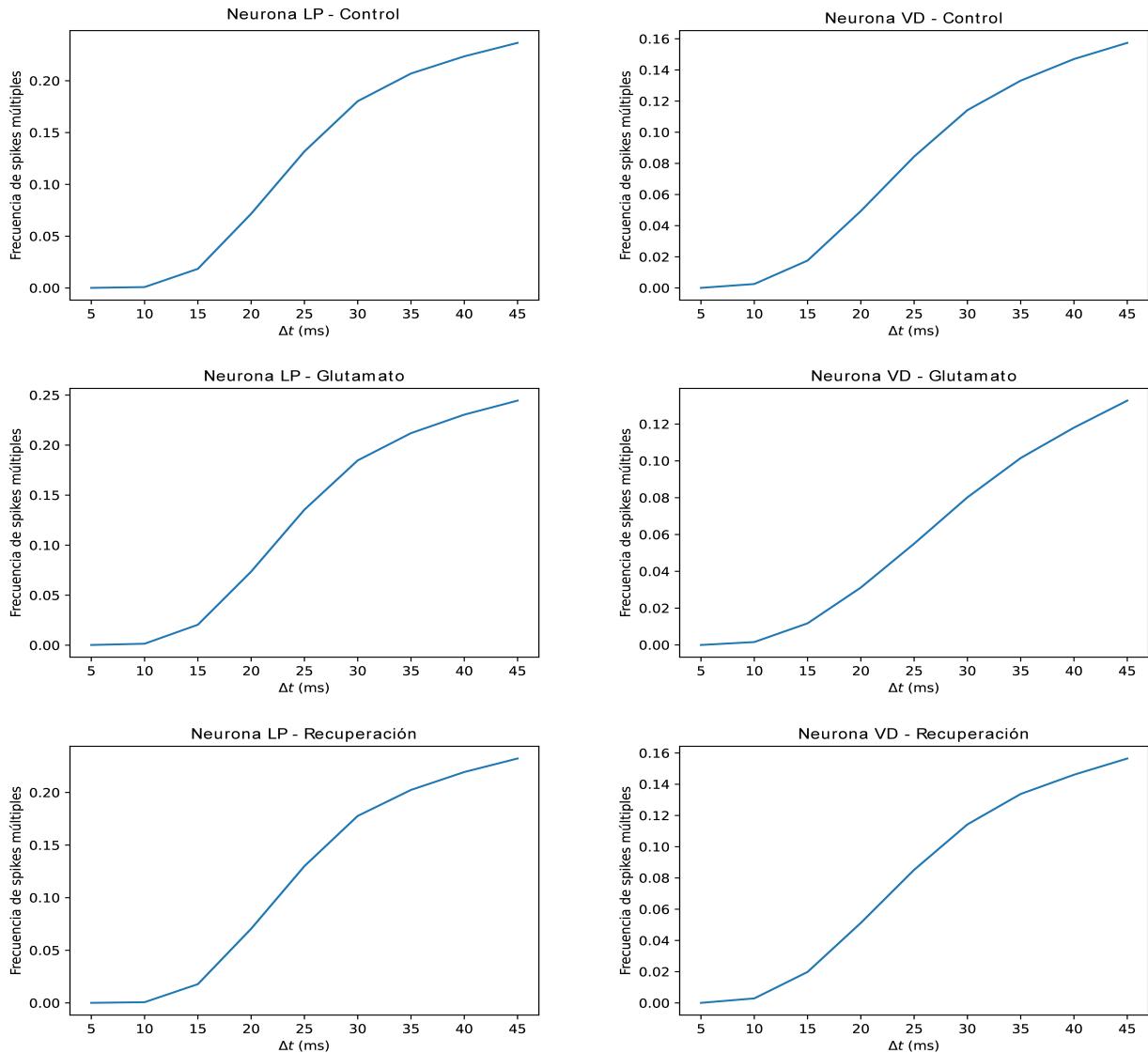


Figura 4: Frecuencia de ventanas que incluían > 1 disparo neuronal en función de Δt . La frecuencia de ventanas con disparos múltiples es prácticamente nula hasta $\Delta t = 10ms$ y permanece razonablemente baja hasta $\Delta t = 15ms$.

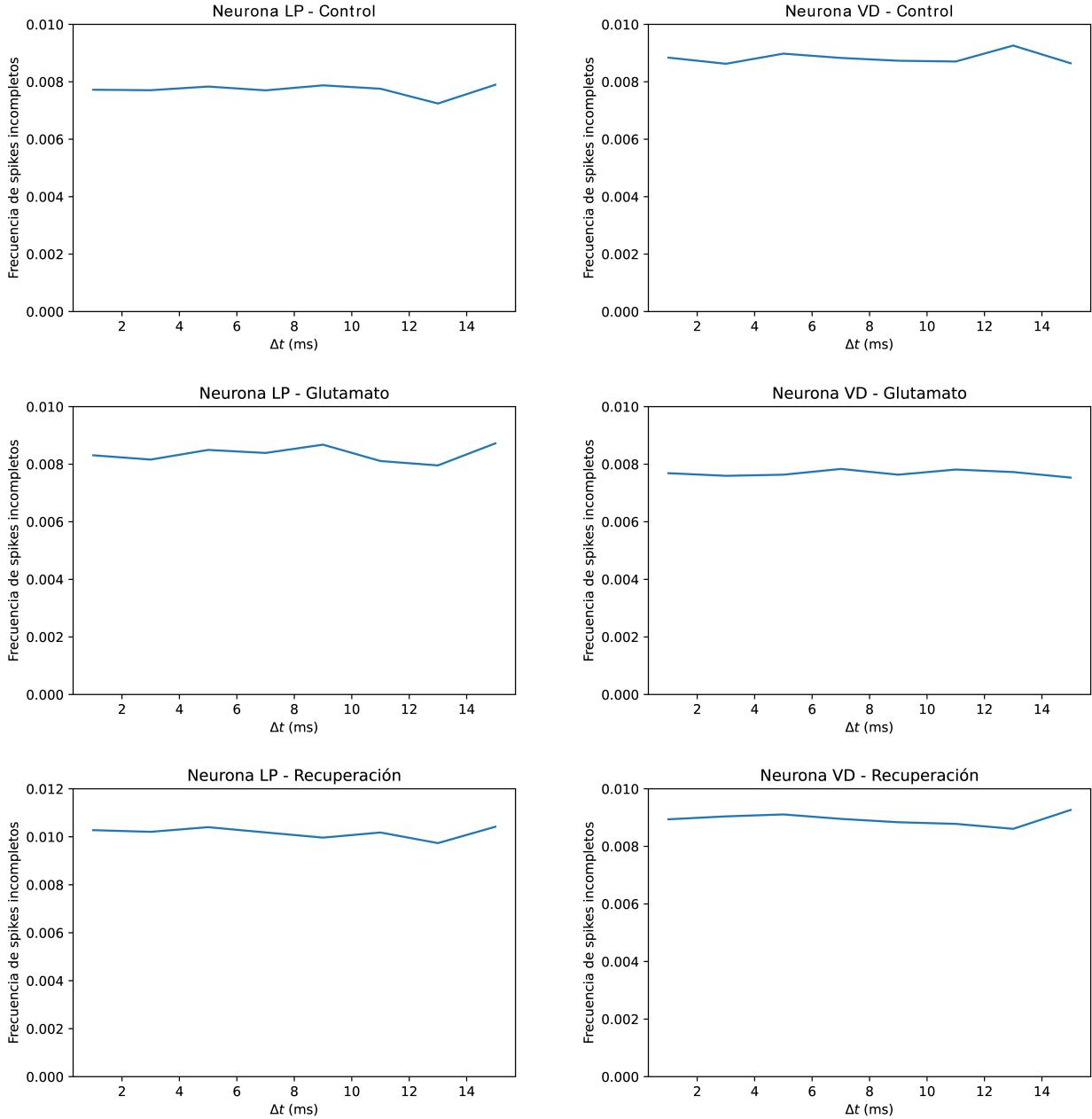


Figura 5: Frecuencia de ventanas que incluían disparos neuronales que ya aparecían en ventanas previas. La frecuencia de ventanas con disparos repetidos es bastante baja para todo Δt y no parece presentar mínimos o máximos concretos.

1.4 Codificación en palabras

Tras el enventanado de las series de datos, fuere con la codificación binaria o con SAX (Secciones 1.5), se generaron las series $V^{n*}(w)$ codificadas en palabras de n bits de la siguiente forma:

$$V^{n*}(w_i) = (V^* w_i, \dots, V^* w_{i+n}) \quad (1.5)$$

Donde la nueva serie en la ventana i contiene los valores de la serie original en las ventanas i a $i + n$.

1.5 Codificación por SAX

La codificación por Symbolic Aggregate Approximation (SAX) consiste en 2 etapas:

1. Piecewise Aggregate Approximation (PAA): enventanado de las series de datos $V_z(t)$ con tamaño de ventana Δt para dar lugar a la nueva serie $V_{PAA}^*(w)$, tal que su valor en cada ventana es el promedio de los valores de la serie original:

$$V_{PAA}^*(w_i) = \frac{1}{n} \sum_{t=t_i}^{t_i+\Delta t} V_z(t) \quad (1.6)$$

2. Symbolic codification: discretización de $V_{PAA}^*(w)$ en $V_{SAX}^*(w)$. El método SAX asume que $V(t) \sim N(\mu, \sigma^2)$, por lo que la serie normalizada $V_z(t) \sim N(0, 1)$. Así, $V_{PAA}^*(w)$ se discretiza en k valores de acuerdo a los cuantiles de una $N(0, 1)$ entre $[0, 1]$:

$$V_{SAX}^*(w_i) = \begin{cases} 0, & V_{PAA}^*(w_i) \leq \Phi^{-1}(1/n) \\ \dots \\ n-1, & V_{PAA}^*(w_i) \leq \Phi^{-1}(n/n) \end{cases} \quad (1.7)$$

La elección del rango de valores de Δt y k considerados se detalla en el Anexo 6.1. La Figura 6 muestra un extracto de las gráficas de transformación PAA y codificación SAX incluidas en el Anexo. En ella se aprecia como la codificación SAX refleja los picos y valles de la señal original.

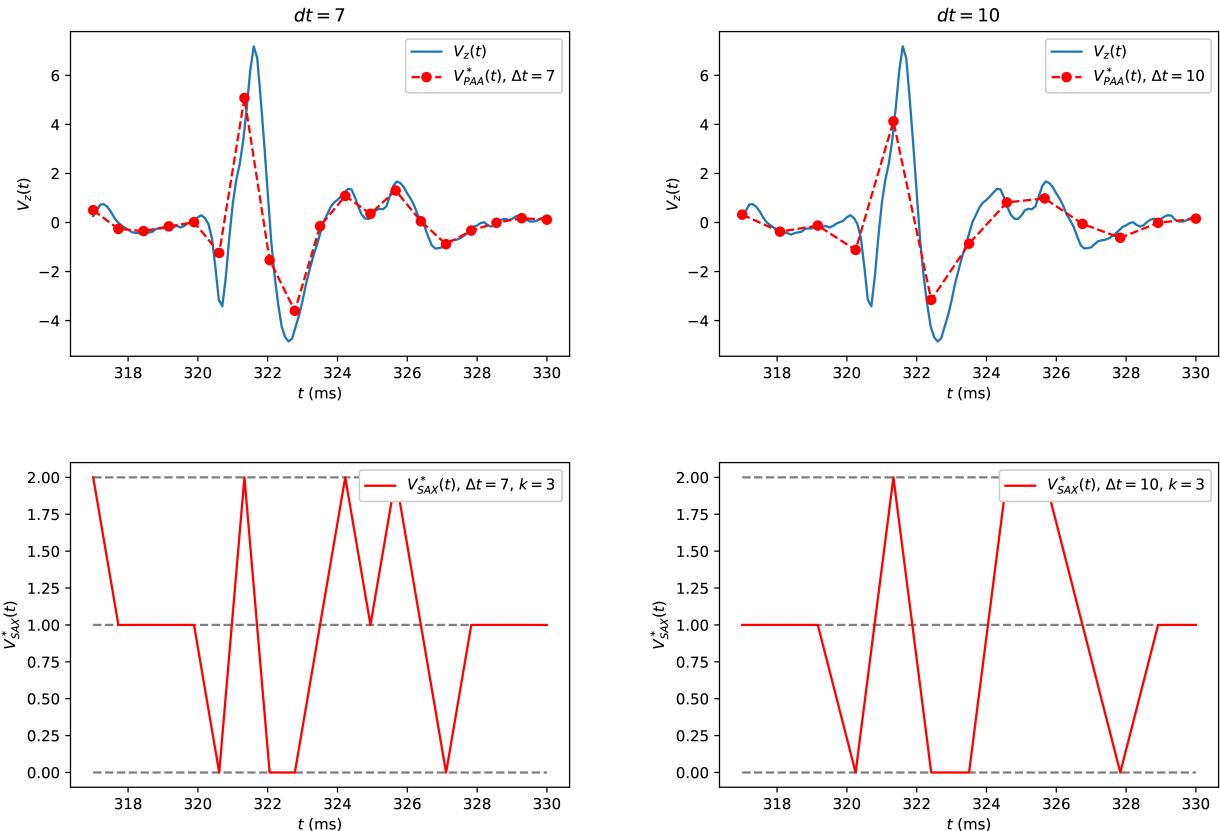


Figura 6: PAA (arriba) y codificación SAX (abajo) $k = 3$ de un intervalo de la serie LP - Control que incluye un disparo neuronal para $\Delta t = 0.7$ y 1. La codificación SAX refleja correctamente los picos y depresiones del disparo neuronal.

1.6 Codificación por patrones de orden

La codificación por patrones de orden también comprende 2 etapas:

1. PAA: equivalente a lo descrito en la Sección 1.5. El rango de valores seleccionado también corresponde al del Anexo 6.1.
2. Codificación por patrones de orden: similar a la codificación por palabras descrita en la Sección 1.4, donde se genera la nueva serie $V_{OP}^{k*}(w)$ de orden k de la siguiente forma:

$$V_{OP}^{k*}(w_i) = \left(OP[V_{PAA}^*(w_i), V_{PAA}^*(w_i + 1)], \dots, OP[V_{PAA}^*(w_i + k - 1), V_{PAA}^*(w_i + k)] \right) \quad (1.8)$$

$$OP[V(w_1), V(w_2)] = \begin{cases} 0, & |V(w_1) - V(w_2)| < \epsilon \\ -1, & V(w_1) - V(w_2) < -\epsilon \\ 1, & V(w_1) - V(w_2) > \epsilon \end{cases} \quad (1.9)$$

donde ϵ es un umbral de tolerancia.

Es decir, la nueva serie en la ventana i contiene los resultados de la operación $OP[\dots]$ entre los pares de ventanas consecutivas que hay desde i hasta $i + k$. A su vez, el resultado de la operación $OP[\dots]$ es 0 si los valores en las ventanas no difieren más de la tolerancia ϵ , y -1 o 1 según si la serie desciende o asciende de una ventana a la otra.

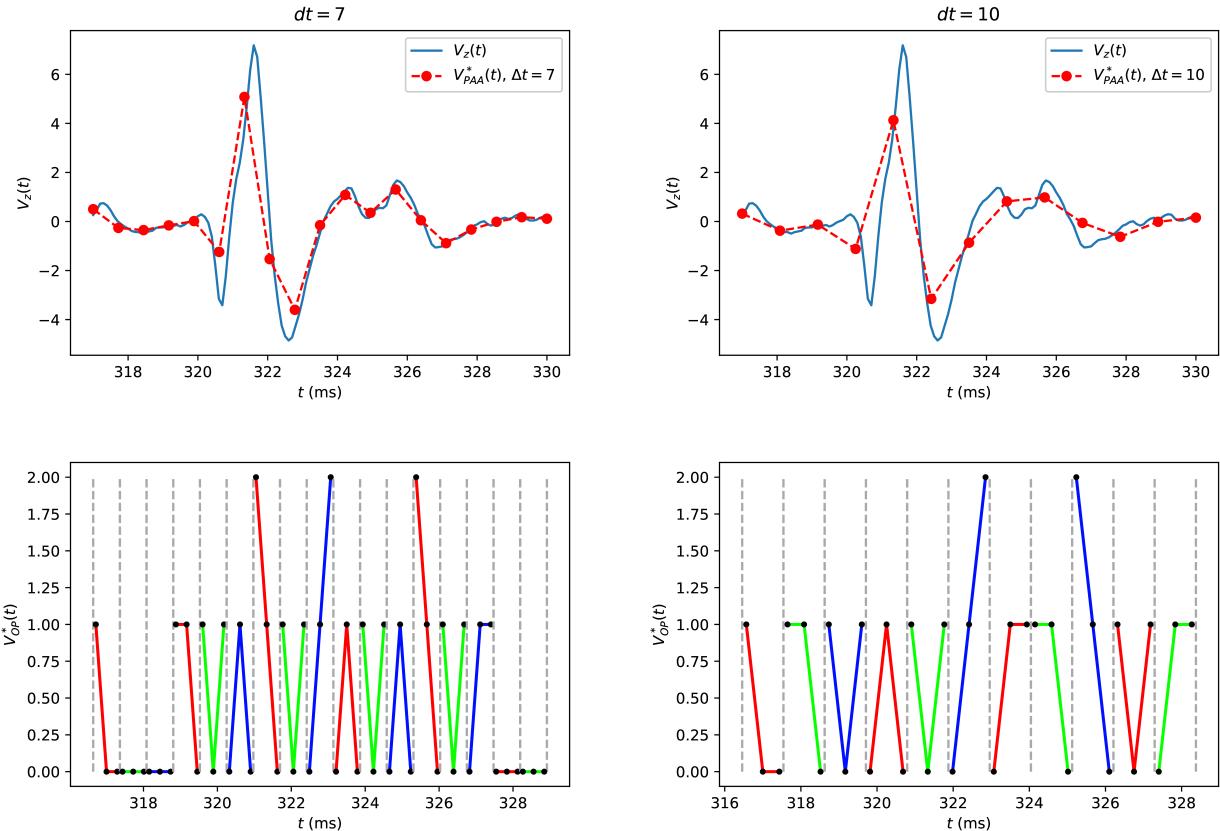


Figura 7: PAA (arriba) y codificación OP (abajo) $k = 3$ de un intervalo de la serie LP - Control que incluye un disparo neuronal para $\Delta t = 0.7$ y 1. La codificación OP refleja correctamente los picos y depresiones del disparo neuronal.

La elección del rango de valores de k considerados se detalla en el Anexo 6.2. La Figura 7 muestra un extracto de las gráficas de transformación PAA y codificación OP incluidas en el Anexo. En ella se aprecia como la codificación OP refleja la señal original.

2 IM y IT con codificación binaria

La Figura 8 muestra la información mutua (IM) entre las neuronas LP y VD para las series Control, Glutamato y Recuperación con la codificación binaria. En las tres series se observa un crecimiento monótono con Δt y una "esquina", a partir de la cual disminuye el ritmo de crecimiento, en un Δt que decrece con el tamaño de palabra. Así mismo, la IM entre las series Control y Recuperación es similar, pero es menor en el caso de Glutamato.

En base a estas observaciones se plantean las siguientes conclusiones:

- Existe un cierto desfase entre ambas neuronas que causa el aumento de MI al aumentar Δt .
- La "esquina" que se observa en las gráficas podría producirse porque una vez la serie incorpora información temporal en un entorno suficientemente grande, la ganancia de IM se vuelve menor. Esto explicaría porque aparece a menores Δt al aumentar el tamaño de palabra, un mayor tamaño de palabra da lugar a incluir información temporal en un intervalo más amplio.
- La inyección de glutamato en el modo de operación utilizado en estas series reduce la comunicación entre las neuronas. Esto podría deberse a una pérdida de relación entre las señales o a un incremento de su desfase.

La Figura 9 muestra la transferencia de información (IT ; entropía relativa) entre las neuronas $LP \rightarrow VD$ y $VD \rightarrow LP$ para las 3 series con codificación binaria. Se vuelve a observar un crecimiento monótono y una "esquina" que aparece en Δt menor con el tamaño de palabra. De nuevo, la serie de Glutamato tiene menores ITs tanto en $LP \rightarrow VD$ como en $VD \rightarrow LP$. Por otro lado, la IT entre $LP \rightarrow VD$ es significativamente mayor que entre $VD \rightarrow LP$ en todos los casos.

Estos resultados reafirman las conclusiones planteadas en el estudio de la MI y sugieren una nueva:

- La IT entre $LP \rightarrow VD$ es mayor que entre $VD \rightarrow LP$ porque la neurona LP se conecta hacia la VD. Debido a esa dirección causal, la información de LP se transfiere más fácilmente a LP que al revés (VD puede sufrir alteraciones por otros factores, como otras neuronas, que no son explicables únicamente por LP)

Finalmente, la Figura 10 muestra el comportamiento de la MI con valores de Δt mucho mayores al intervalo considerado. En estos valores hay una gran frecuencia de varios disparos incluidos en la misma ventana y se produce una saturación y posterior decrecimiento de la IM que no responde a una verdadera pérdida de transmisión entre las neuronas, sino a la pérdida de información en la codificación. Esta saturación también se observaba en las figuras anteriores con $\Delta t > 15$.

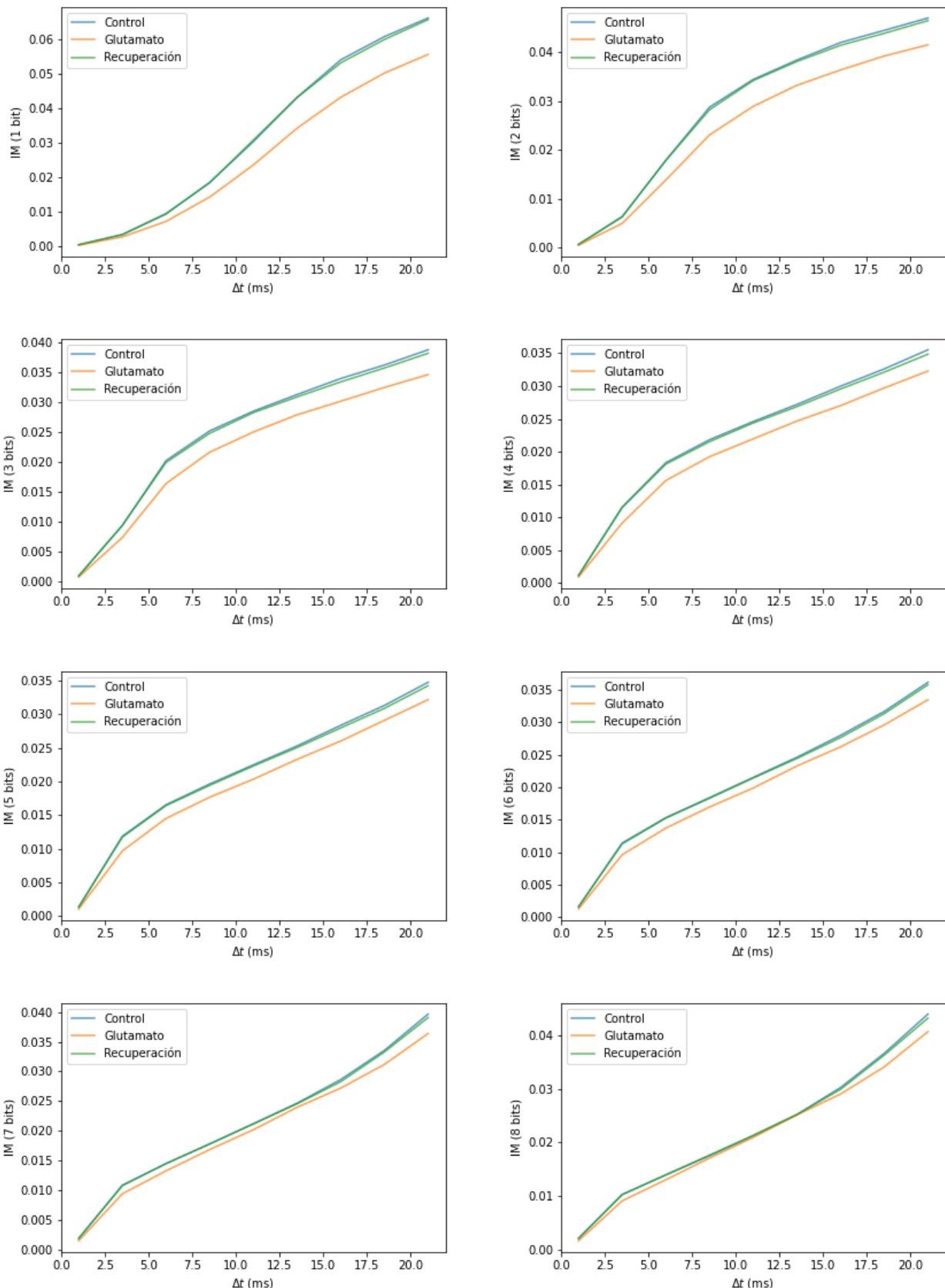


Figura 8: IM entre las neuronas LP y VD para las series Control (C), Glutamato (G) y Recuperación (R) para un Δt de $[1, 20]$ y un tamaño de palabra $n = [1, 8]$ con codificación binaria. En las tres series la IM crece monótonamente con Δt y presenta un pico en un valor de Δt que decrece con el tamaño de palabra.

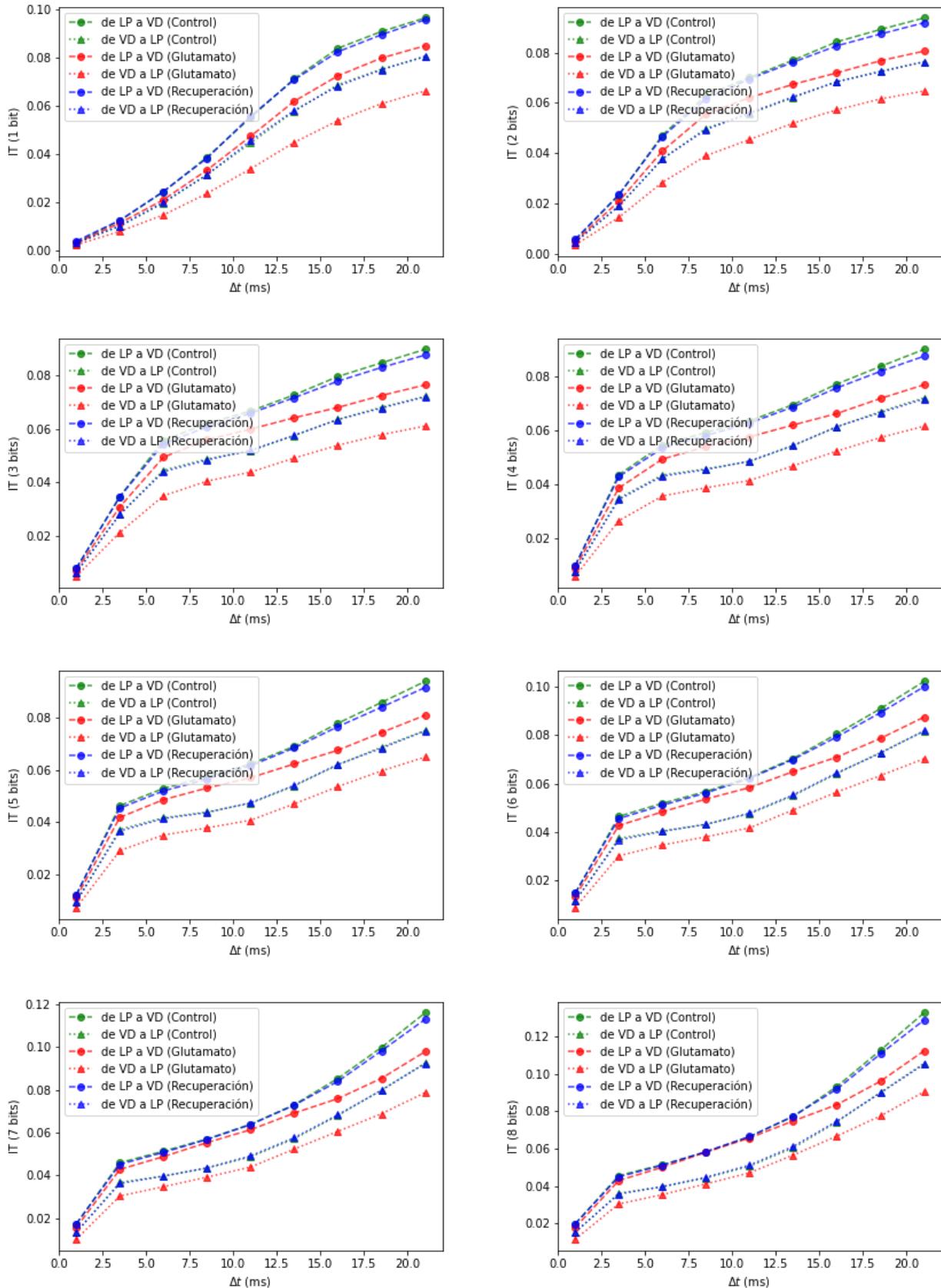


Figura 9: IT entre las neuronas LP \rightarrow VD y VD \rightarrow LP para las series Control (C), Glutamato (G) y Recuperación (R) para un Δt de $[1, 20]$ y un tamaño de palabra $n = [1, 8]$ con codificación binaria. En las tres series la IT crece monótonamente con Δt y presenta un pico en un valor de Δt que decrece con el tamaño de palabra.

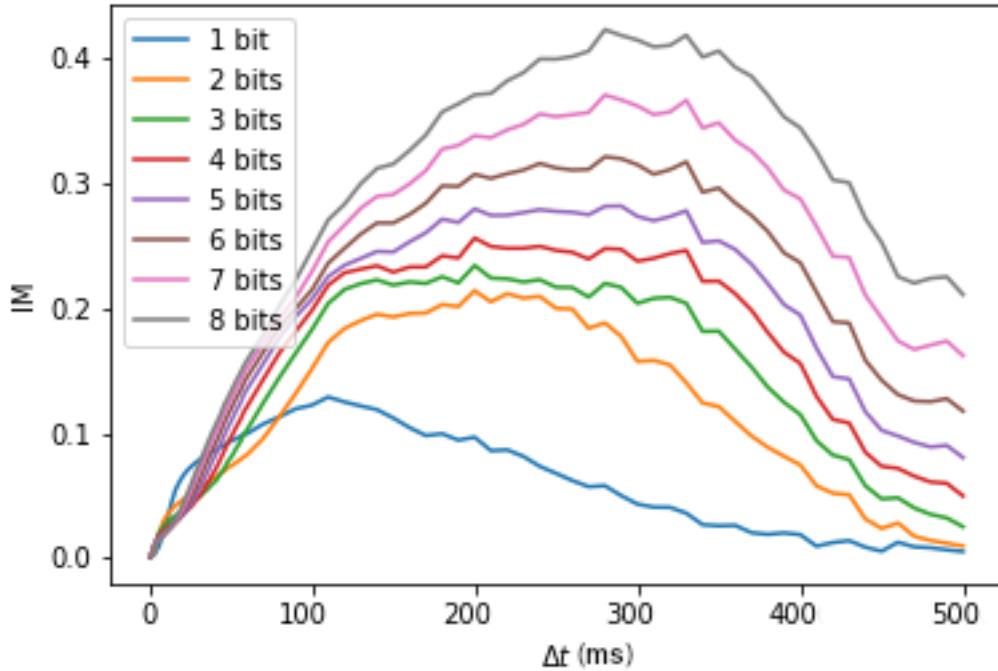


Figura 10: IM entre las neuronas LP y VD para la serie Control para Δt de $[1, 500]$. Con valores de Δt tan altos la probabilidad de varios disparos en la misma ventana es muy alta y la IM desciende porque se pierde la información en la codificación.

3 IM y IT con codificación SAX

La Figura 11 muestra la IM entre las neuronas LP y VD para las series Control, Glutamato y Recuperación con la codificación SAX. En las tres series se observa que a medida que con el aumento de n se pasa de un crecimiento con Δt a un decrecimiento, estando el punto de transición a Δt más bajo cuanto mayor es n . Así mismo, la IM sigue siendo mínima para la serie de Glutamato, y la segunda menor corresponde a Control.

En base a estas observaciones se vuelven a reforzar conclusiones previas:

- El cambio entre crecimiento y decrecimiento con Δt es similar a la "esquina" que se observaba en la codificación binaria e indica un máximo de información al considerar un cierto intervalo temporal, que podría corresponder al desfase con que la señal de LP alcanza VD. De nuevo, el Δt necesario es menor cuanto mayor es n porque se incluye más información temporal.
- La inyección de glutamato perturba la transferencia de información entre las neuronas.

La Figura 12 muestra la IT entre las neuronas $LP \rightarrow VD$ y $VD \rightarrow LP$ para las 3 series con la codificación SAX. La transferencia entre $LP \rightarrow VD$ muestra un crecimiento monótono con Δt para todo n mientras que entre $VD \rightarrow LP$ se reduce con bajo n o se mantiene constante con alto n

De nuevo, estos resultados reafirman las conclusiones anteriores y la mejor transmisión $LP \rightarrow VD$ por la dirección causal entre ellas.

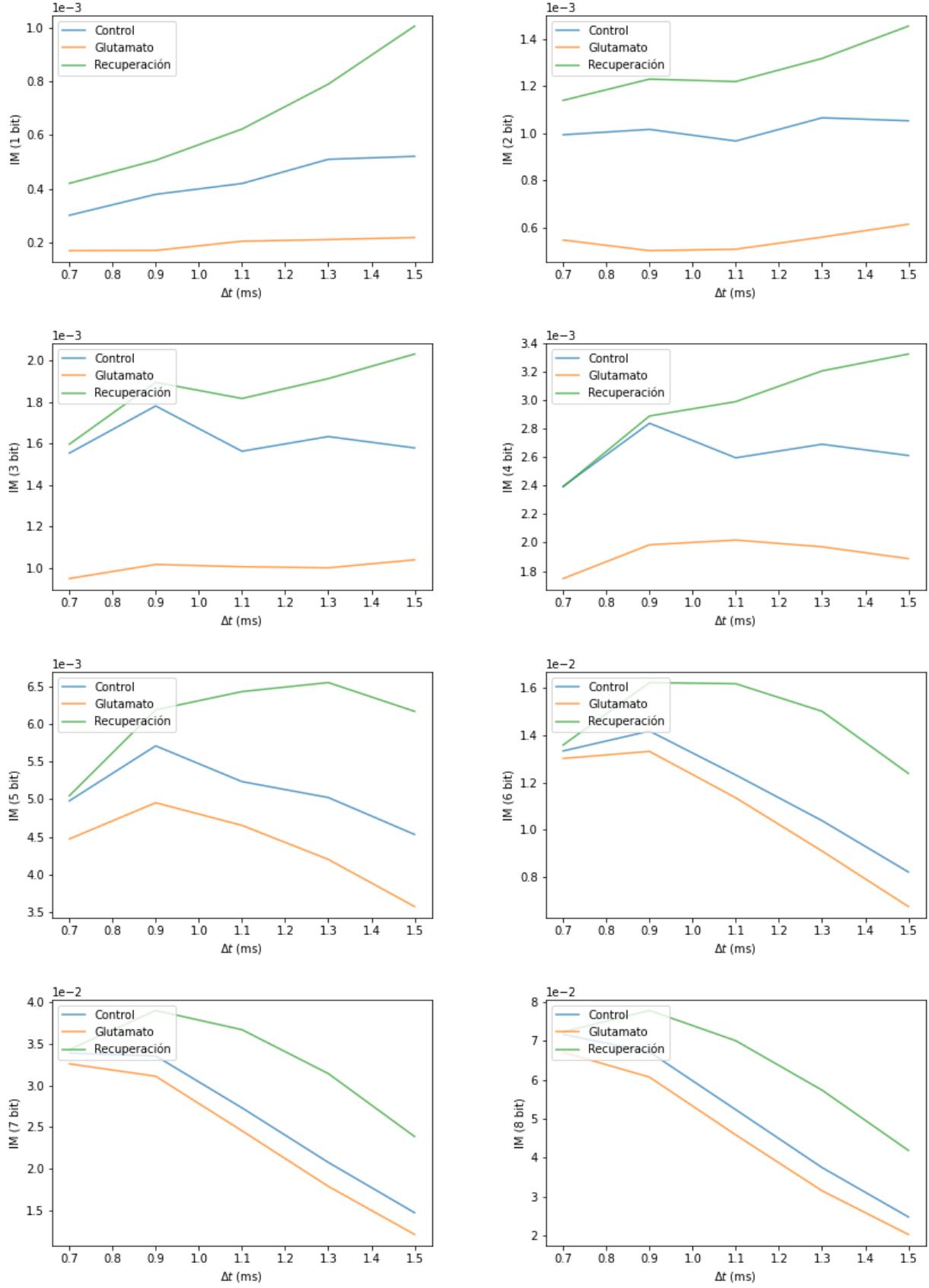


Figura 11: IM entre las neuronas LP y VD para las series Control (C), Glutamato (G) y Recuperación (R) para un Δt de $[0.7, 1.5]$ y un tamaño de palabra $n = [1, 8]$ con codificación SAX. En las tres series la IM alcanza un máximo con Δt que está a valores menores cuanto mayor es el tamaño de palabra.

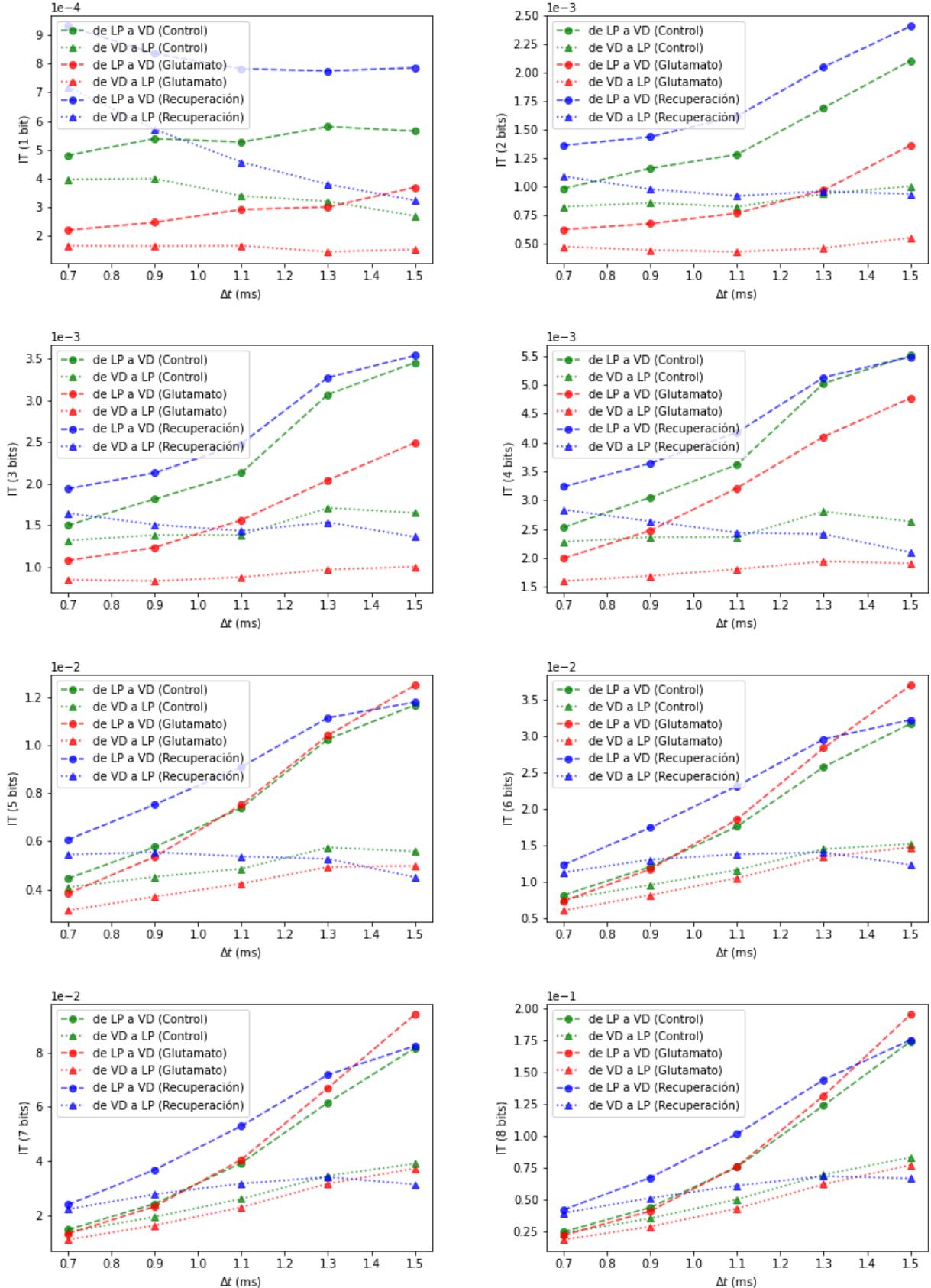


Figura 12: IT entre las neuronas $LP \rightarrow VD$ y $VD \rightarrow LP$ para las series Control (C), Glutamato (G) y Recuperación (R) para un Δt de $[0.7, 1.5]$ y un tamaño de palabra $n = [1, 8]$ con codificación SAX. En las tres series la IT crece monótonamente con Δt para $LP \rightarrow VD$ y se mantiene o se reduce en el caso contrario.

4 IM y IT con codificación OP

La Figura 13 muestra la IM entre las neuronas LP y VD para las series Control, Glutamato y Recuperación con la codificación OP. En las tres series se vuelve a observar que a medida que aumenta n se pasa a un decrecimiento con Δt . Así mismo, la IM sigue siendo mínima para la serie de Glutamato.

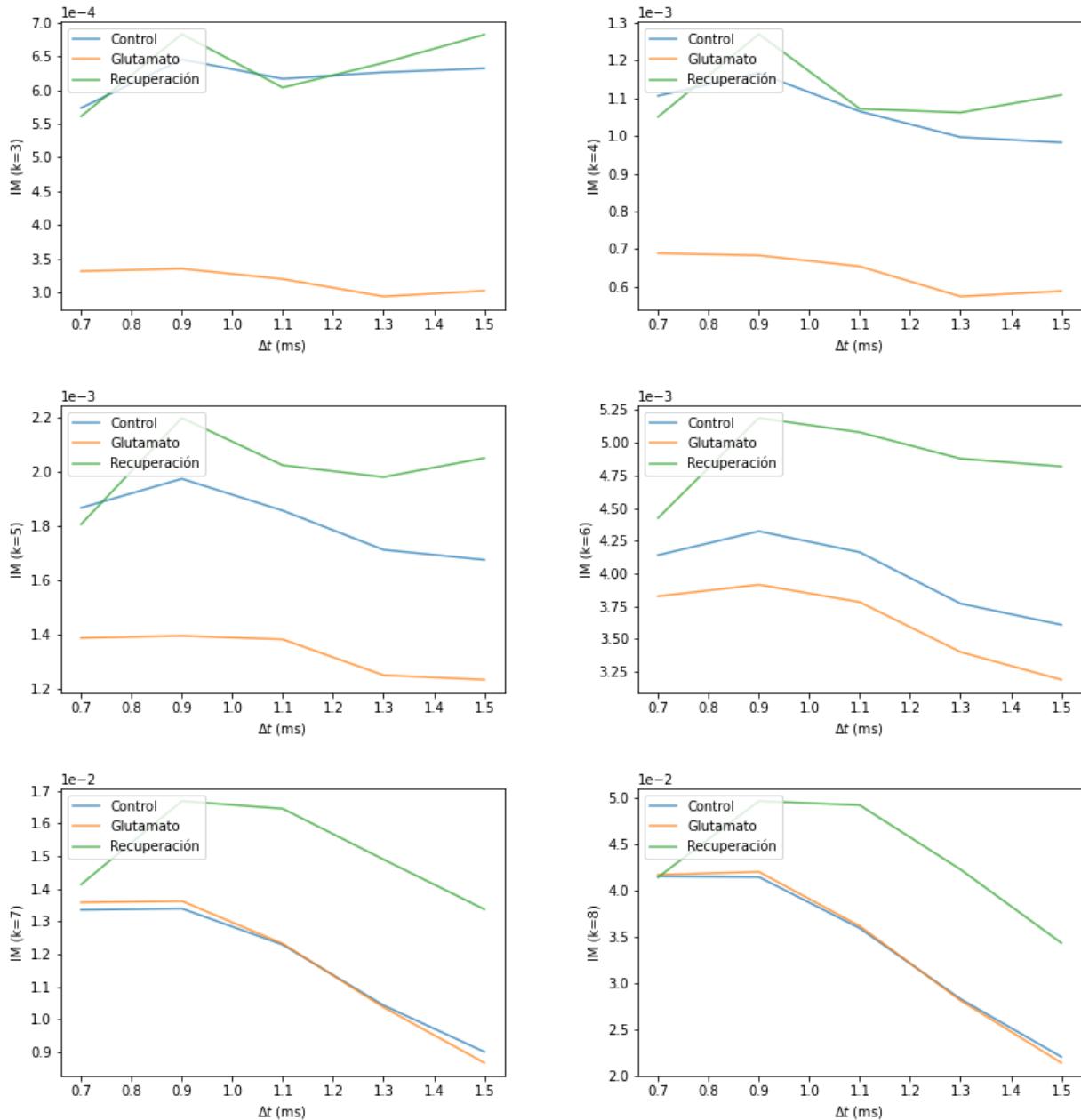


Figura 13: IM entre las neuronas LP y VD para las series Control (C), Glutamato (G) y Recuperación (R) para un Δt de $[0.7, 1.5]$ y un tamaño de palabra $n = [1, 8]$ con codificación OP. En las tres series la IM alcanza un máximo con Δt que está a valores menores cuanto mayor es el tamaño de palabra.

La Figura 14 muestra la IT entre las neuronas $LP \rightarrow VD$ y $VD \rightarrow LP$ para las 3 series con la codificación OP. La transferencia entre $LP \rightarrow VD$ y entre $VD \rightarrow LP$ tiene un comportamiento similar a la IM .

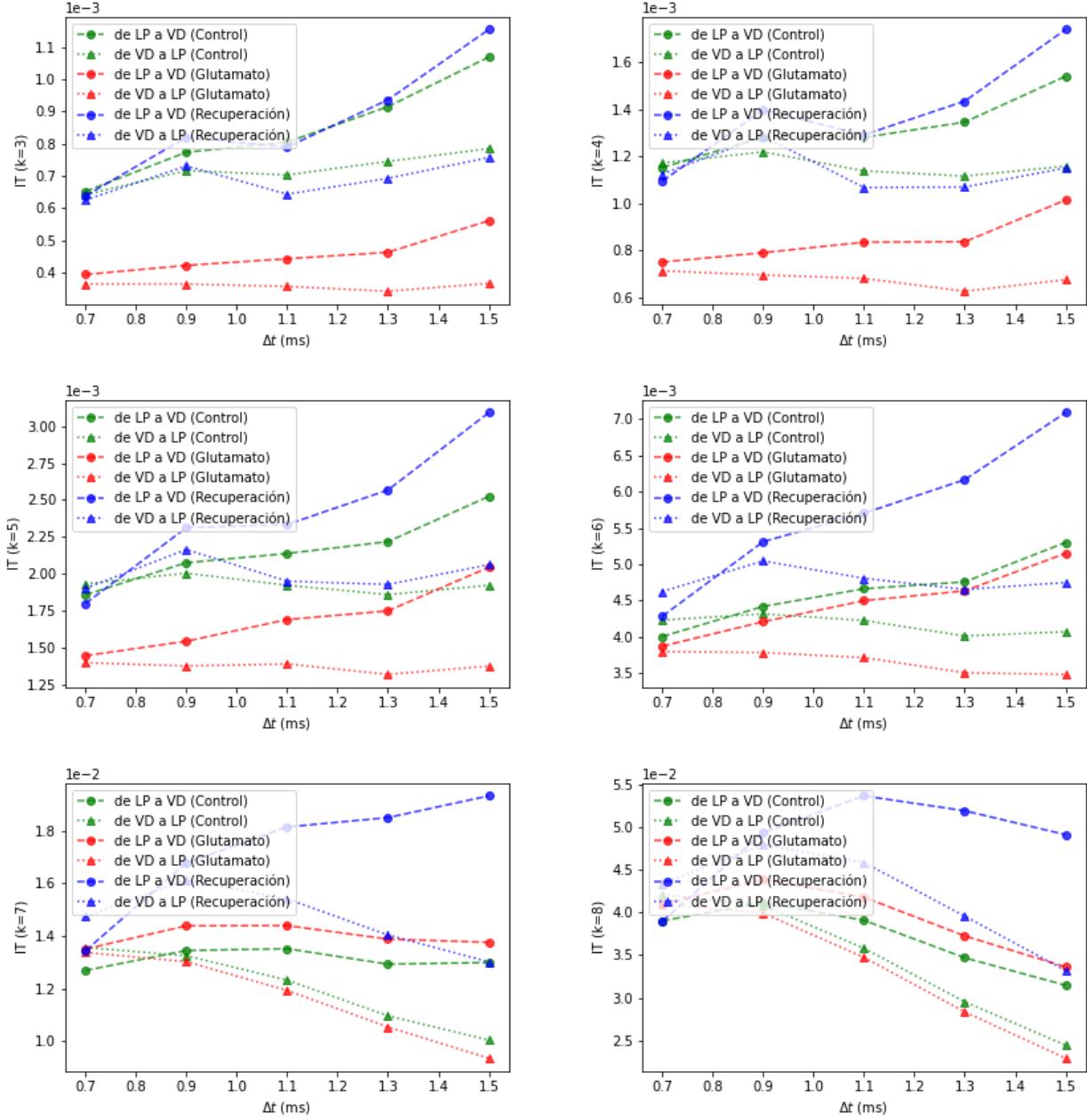


Figura 14: IT entre las neuronas $LP \rightarrow VD$ y $VD \rightarrow LP$ para las series Control (C), Glutamato (G) y Recuperación (R) para un Δt de $[0.7, 1.5]$ y un tamaño de palabra $n = [1, 8]$ con codificación OP.

5 Conclusiones

1. Existe una transferencia de información entre las neuronas LP y VD que alcanza un máximo en valor o en tasa de crecimiento (según el tipo de codificación) con un cierto intervalo temporal que podría corresponder al desfase de comunicación entre ambas.
2. La inyección de glutamato dificulta o retarda la comunicación entre las neuronas LP y VD.
3. La transferencia de información IT es mayor de $LP \rightarrow VD$ porque es la dirección de influencia causal (LP se conecta hacia VD).

6 Anexo

6.1 Parámetros de codificación SAX

La 15 muestra el resultado de la transformación PAA para diferentes valores de Δt . Valores de $\Delta t < 7$ dan lugar a una $V_{PAA}^*(w_i)$ que apenas difiere de la $V_z(t)$ original, mientras que valores de $\Delta t > 15$ dan lugar a una $V_{PAA}^*(w_i)$ prácticamente plana. En consecuencia se consideró un intervalo de valores para Δt de $[7, 15]ms$.

Por otro lado, las Figuras 16 y 17 muestran el resultado de la codificación SAX con $k = 3$ y $k = 4$ respectivamente para diferentes valores de Δt . La discretización con $k = 4$ no reproduce la serie original de forma mucho más precisa que $k = 3$ por lo que se escogió utilizar éste último valor.

6.2 Parámetros de codificación OP

Las figuras 18 y 19 muestran el resultado de la codificación OP con $k = 3$ y $k = 4$ respectivamente, y con $\epsilon = 0.5$ para diferentes valores de Δt en la transformación PAA. Los patrones de orden k tienen un comportamiento y coste computacional análogo a la codificación por palabras, por lo que se consideraron valores de k de $[3, 8]$.

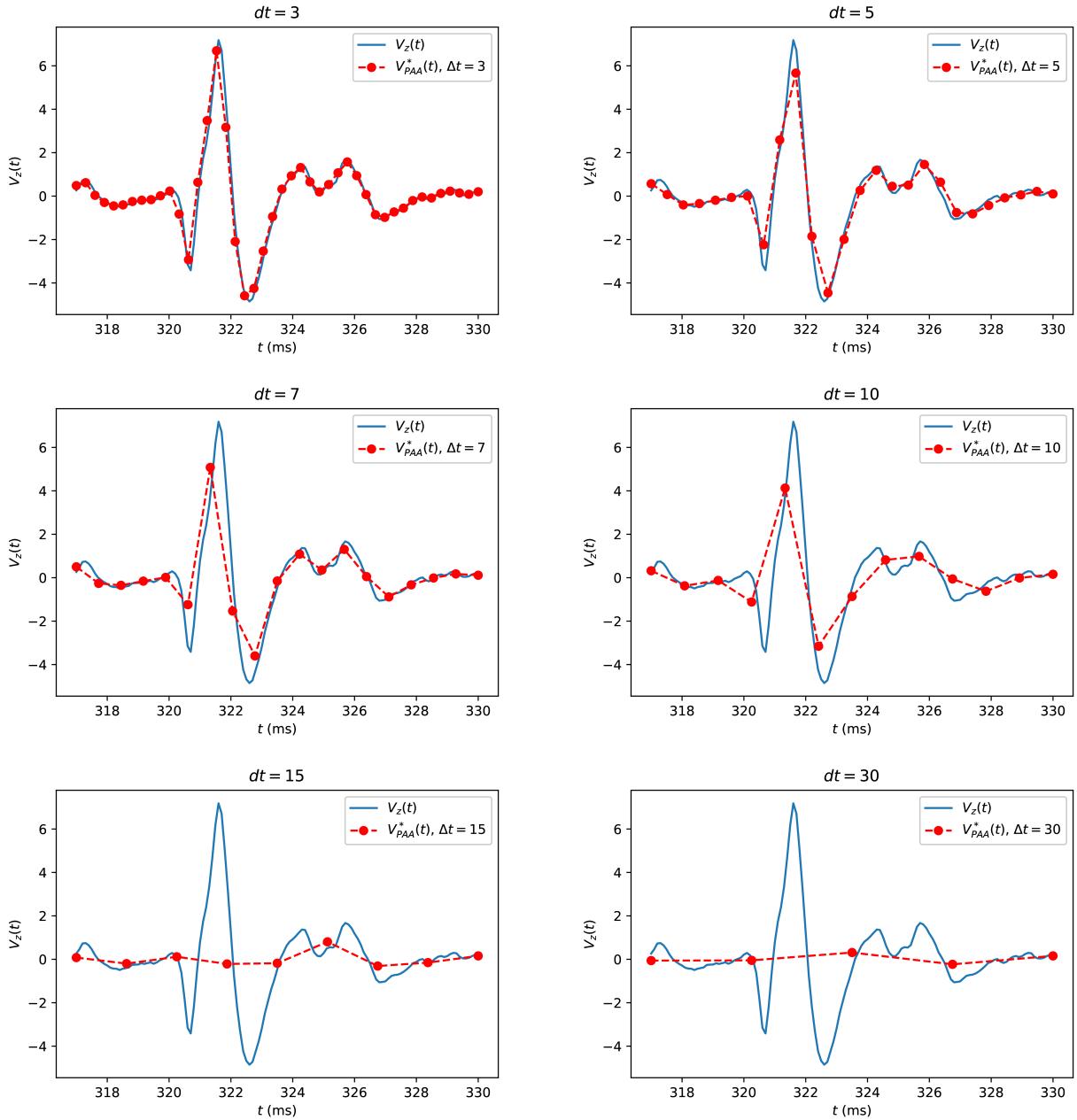


Figura 15: PAA de un intervalo de la serie LP - Control que contiene un disparo neuronal en función del tamaño de ventana Δt . $\Delta t < 7$ da lugar a series prácticamente iguales a la original, mientras que $\Delta t > 15$ da lugar a series demasiado planas.

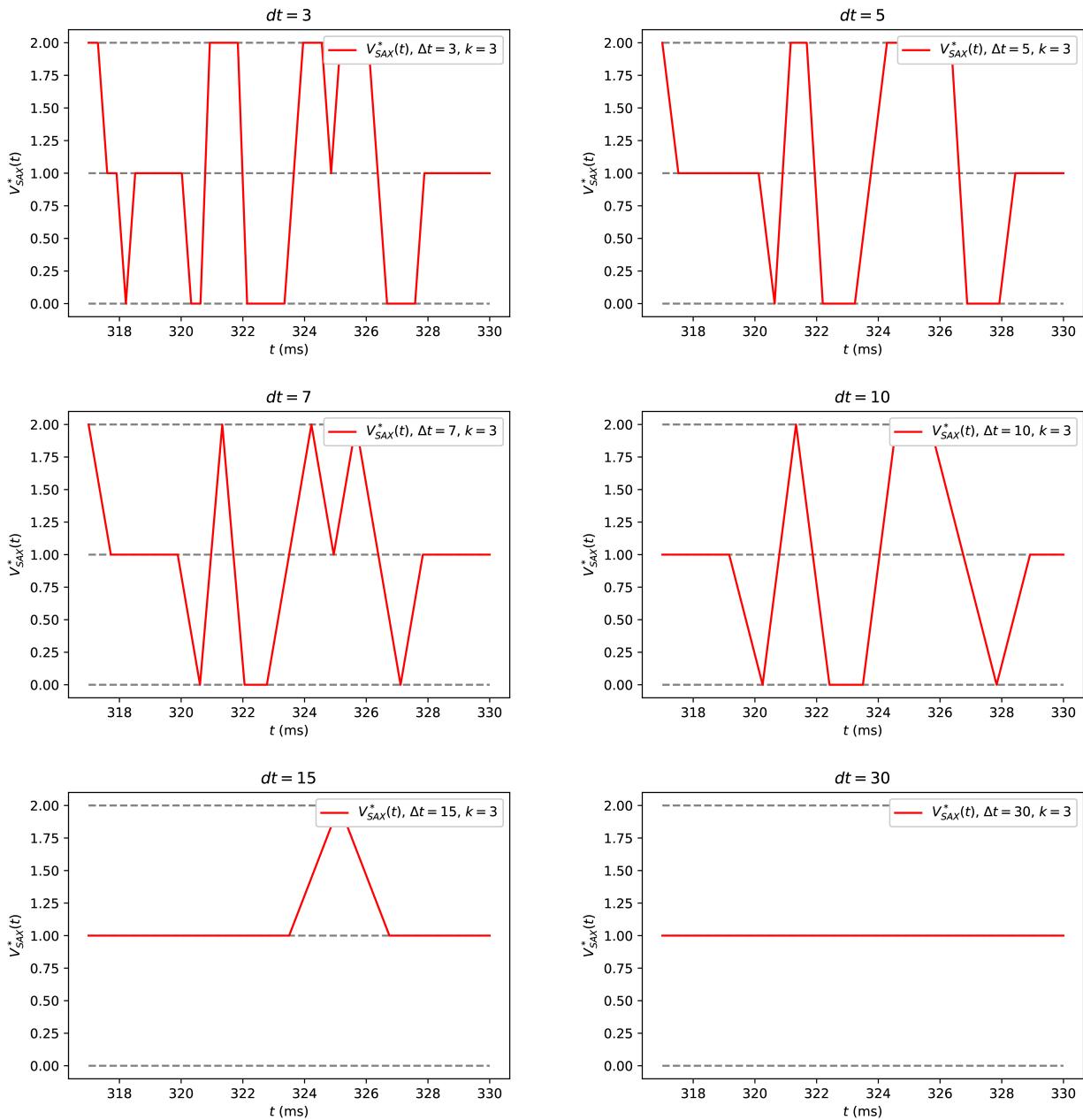


Figura 16: Codificación SAX con $k = 3$ de un intervalo de la serie LP - Control que contiene un disparo neuronal en función del tamaño de ventana Δt .

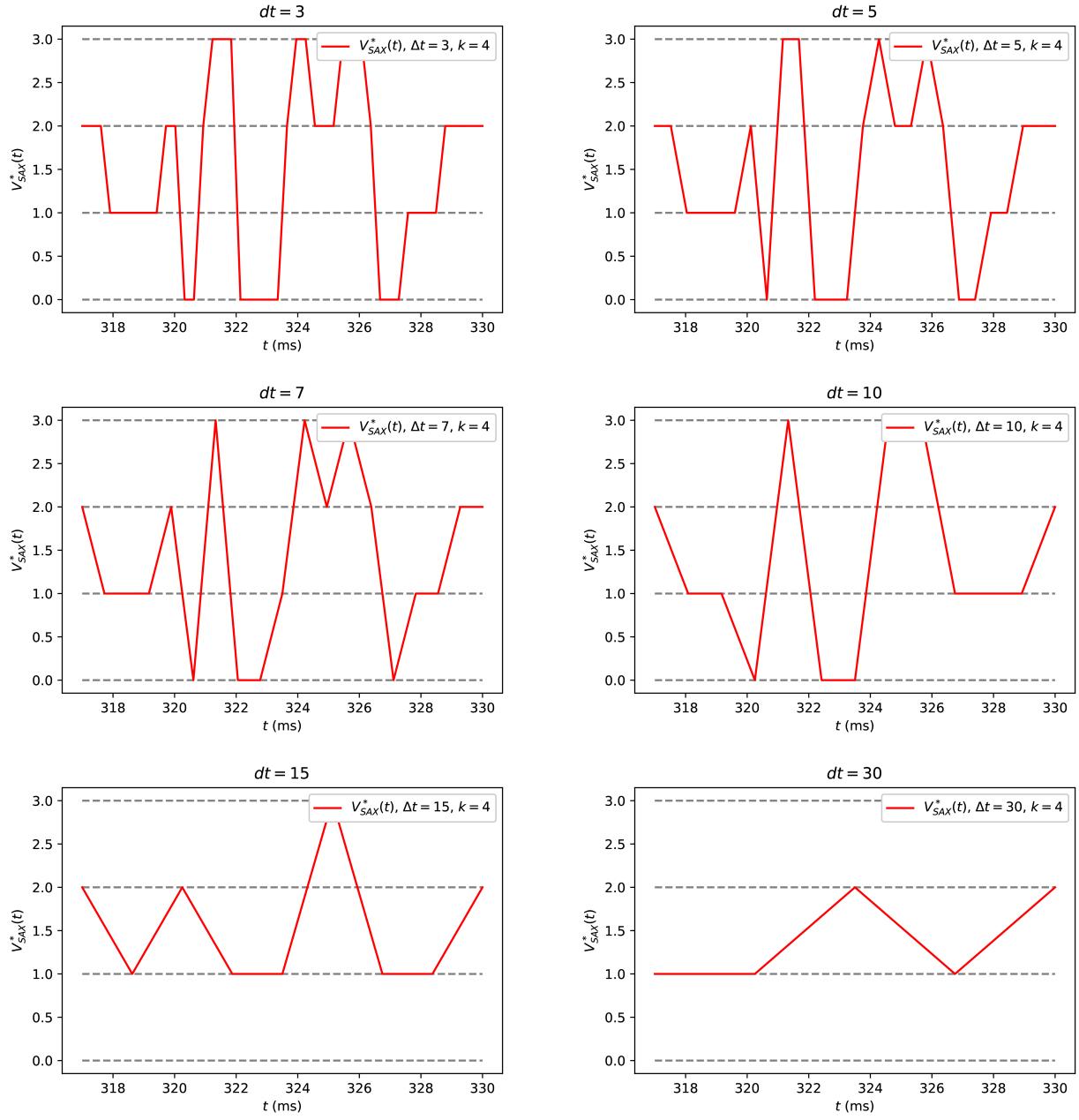


Figura 17: Codificación SAX con $k = 4$ de un intervalo de la serie LP - Control que contiene un disparo neuronal en función del tamaño de ventana Δt .

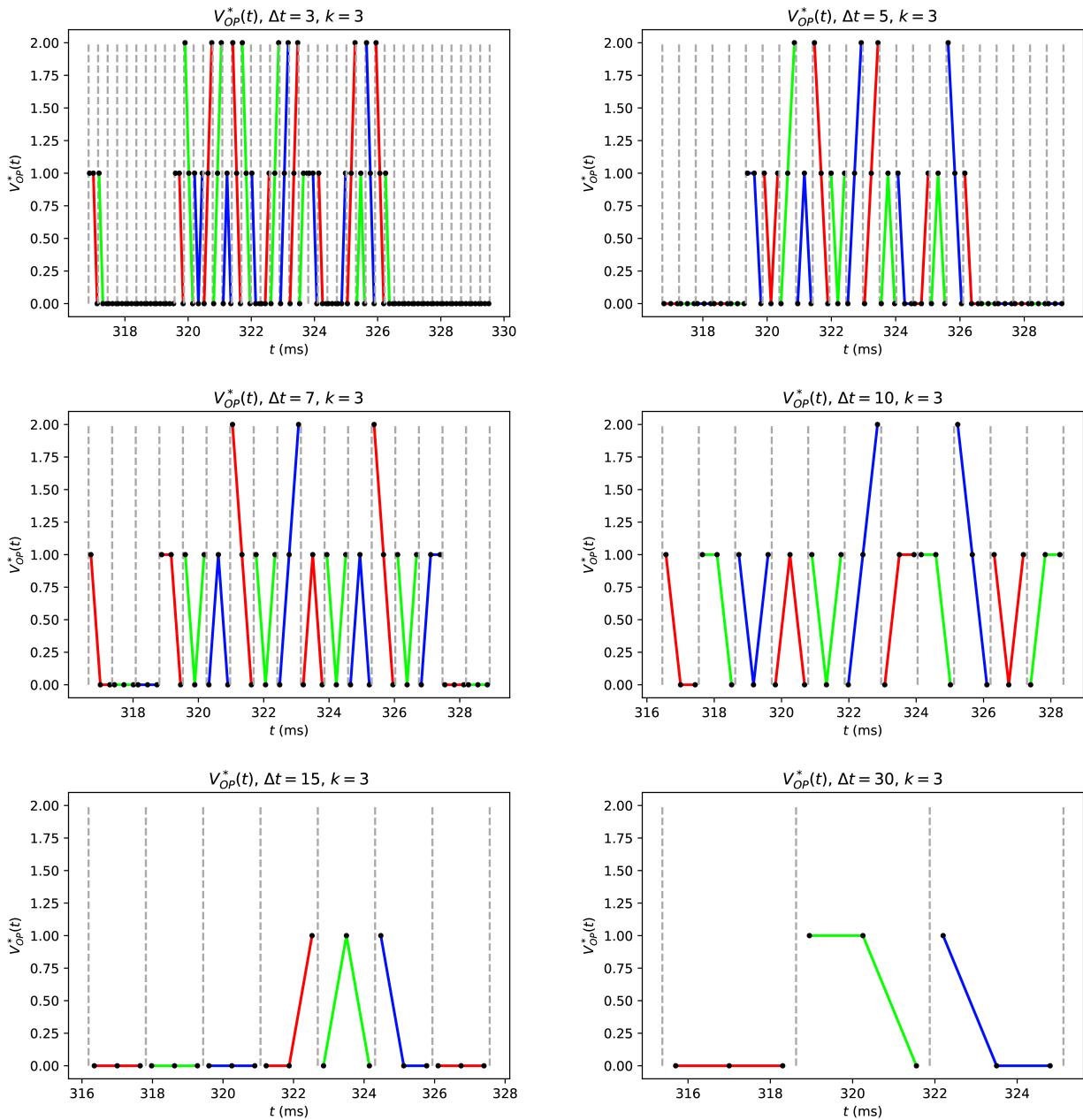


Figura 18: Codificación OP con $k = 3$ de un intervalo de la serie LP - Control que contiene un disparo neuronal en función del tamaño de ventana Δt .

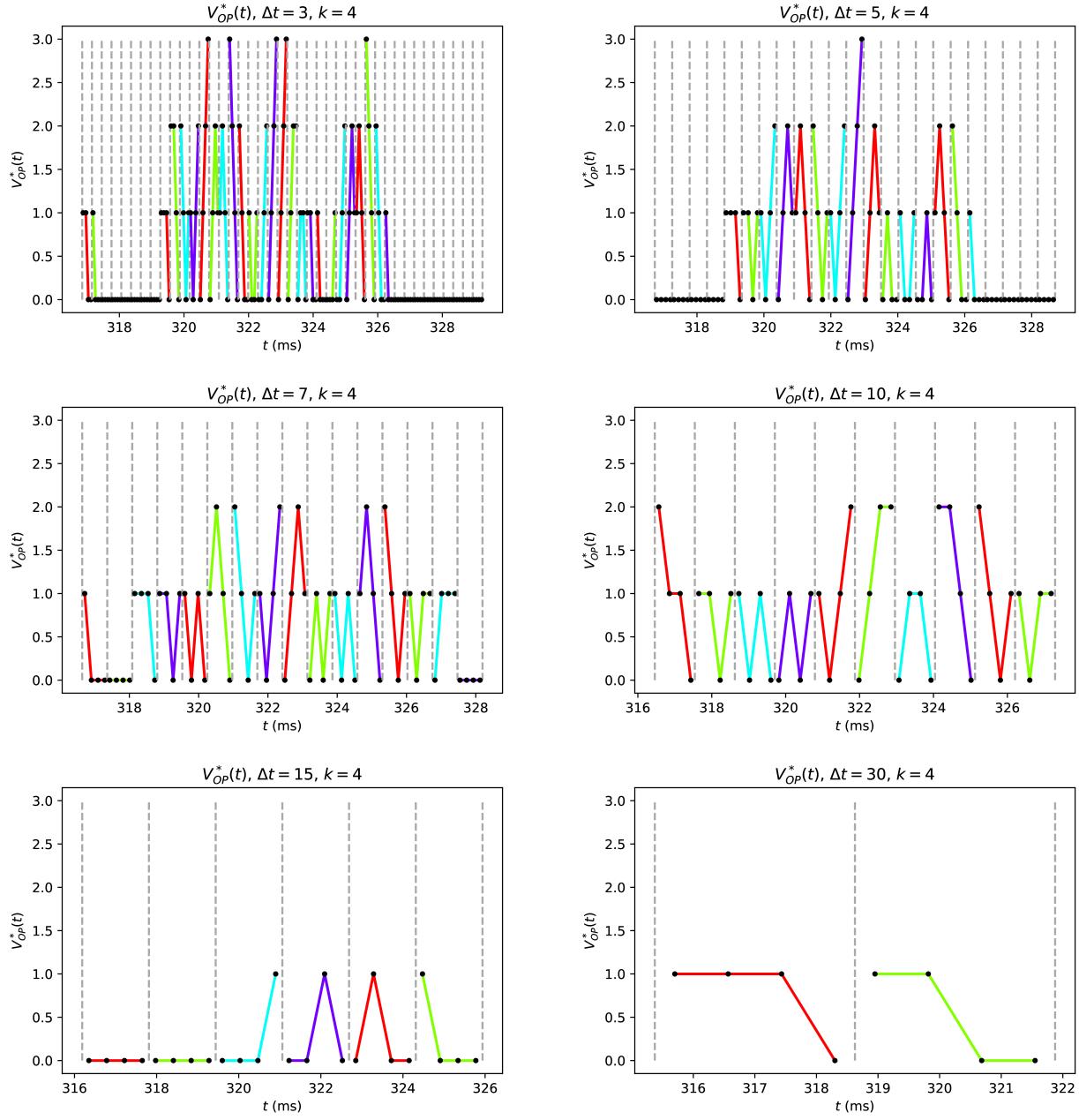


Figura 19: Codificación OP con $k = 4$ de un intervalo de la serie LP - Control que contiene un disparo neuronal en función del tamaño de ventana Δt .