# Estudio acústico del VOT de las oclusivas del español pronunciadas por alumnos sinohablantes de diferentes perfiles lingüísticos[1]

Pengfei Zhai
*Universidad Autónoma de Barcelona*

## 1 Introducción

El concepto de pronunciación hace referencia a la articulación de los segmentos de una lengua, respetando todos los rasgos suprasegmentales como la acentuación, el ritmo y la entonación. En el campo del aprendizaje de segundas lenguas o lenguas no nativas (L2/LE), la buena pronunciación no solamente garantiza que la comunicación oral sea satisfactoria, sino que también genera empatía por parte de la comunidad de la lengua meta. En el ámbito de la pronunciación de lenguas no nativas, es frecuente recurrir al término de interferencia, al considerar que determinados tipos de error se deben a la influencia de la lengua materna del aprendiente. A diferencia de lo que ocurre en la adquisición de la lengua materna (L1), durante el aprendizaje de una L2/LE el aprendiz ya cuenta con su propia fonología que, a su vez, podrá inhibirle a la hora de percibir y/o pronunciar ciertos segmentos de la lengua meta.

Según el modelo *Speech Learning Model* (Flege 1995), cuando existe suficiente diferencia fonética entre un sonido de la L2 y un alófono que le es perceptivamente más cercano de un segmento de la L1, el aprendiente podrá ser capaz de discriminar los dos eficazmente, desarrollando una nueva categoría fonética que especifica y almacena los rasgos del nuevo sonido en la memoria a largo plazo. Contrariamente, si el aprendiente es incapaz de percibir dicha diferencia, un mecanismo denominado "clasificación de equivalencia" (*equivalence classification*) impedirá la formación de la categoría para el sonido no nativo. En este caso, se hipotetiza que el aprendiente pronunciará y percibirá ambos sonidos de manera igual o similar.

En lo que se refiere al aprendizaje del español por parte de hablantes del chino o sinohablantes, la pronunciación de las consonantes oclusivas supone una de las principales dificultades, ya que en el léxico del español existen numerosos pares mínimos que se distinguen solamente por una oclusiva, como en *pala-bala* o *pino-vino*. Asimismo, el error de pronunciación de oclusivas puede alterar totalmente el significado que se quiera transmitir y dificultar considerablemente la comunicación oral.

Contrastando el sistema fonológico del español con el del chino mandarín, parece razonable atribuir la causa de este problema a la interferencia de la L1 del aprendiz. El chino mandarín, que es la variedad del chino más hablada (830 millones de usuarios aproximadamente), y el español contienen las mismas series de oclusivas: bilabiales, dentales[2] y velares. No obstante, las sonoras del español están ausentes en el repertorio fónico del chino mandarín. Mientras las oclusivas castellanas homoorgánicas se distinguen claramente por la sonoridad anticipada, las del chino mandarín se diferencian

[1] El autor agradece sinceramente a Dolors Poch Olivé por sus comentarios sobre la versión original del manuscrito de este trabajo, así como a los evaluadores anónimos por su revisión.

[2] Tradicionalmente, se considera que el punto de articulación de /t/ y /d/ españolas es dental, como en Quilis (1981), Hualde (2005) y Gil (2007). Sin embargo, autores como Martínez Celdrán y Fernández Planas (2007) sostienen que el punto de articulación no es propiamente dental, sino dentoalveolar. La /t/ del chino mandarín y del dialecto wu se describe como dentoalveolar según Lee y Zee (2003) y Chen y Gussenhoven (2015). En cualquier caso, esta cuestión no será tratada específicamente en el presente estudio, pues no incluye datos articulatorios.

por la aspiración. Acústicamente, /p t k/ del español se realizan con VOT[3] positivo corto y /b d g/ con VOT negativo en posición inicial absoluta; en tanto que /p t k/ del chino tienen, igualmente, valor de VOT positivo inferiores a 40 ms. y sus contrapartes /pʰ tʰ kʰ/ tienen VOT próximos a 100 ms.

Aunque a menudo se acostumbra a emplear el concepto *sinohablante* para referirse a 'hablante del chino mandarín', no hay que olvidar que entre el alumnado que comparte la misma lengua materna aún existe relevante heterogeneidad en cuanto a su perfil lingüístico. Es conveniente recordar que la vasta extensión geográfica y el altísimo número de usuarios le brindan al idioma más hablado del mundo abundantes variaciones diatópicas. En lo que se refiere al consonantismo, Lee y Zee (2014) observan que, a diferencia del chino mandarín, 12 dialectos del chino de los 70 investigados cuentan con la diferencia fonológica en sonoridad en su repertorio de oclusivas. Por lo tanto, será interesante examinar el resultado de pronunciación y percepción de alumnos de otros perfiles para contrastarlo con el de hablantes del mandarín. En línea con el modelo *Speech Learning Model*, y dado que tanto las sordas como las sonoras del español forman parte de su lengua nativa, es posible hipotetizar que este conjunto de alumnos sea capaz de establecer una distinción efectiva entre las dos categorías, aventajando así a sus compañeros de chino mandarín en el proceso de aprendizaje.

Entre las mencionadas variedades geográficas del chino que poseen segmentos oclusivos sonoros, destaca el dialecto wu (吳方言), que constituye el segundo dialecto más hablado de China (solo superado por el mandarín), y se utiliza en el sureste del país asiático, incluyendo la zona metropolitana de Shanghái. Contiene tres categorías de oclusivas: sordas no aspiradas (/p t k/), sordas aspiradas (/pʰ tʰ kʰ/) y sonoras (/b d g/) (véase Chen y Gussenhoven 2015). En este sentido, resulta interesante estudiar la pronunciación de sus hablantes, ya que en su inventario fónico está presente el contraste en sonoridad, al igual que en la lengua castellana.

## 2 Metodología

Como intento de describir cualitativa y cuantitativamente la diferencia en la pronunciación de las oclusivas del español por alumnos sinohablantes de distintos perfiles, se pretende medir el VOT de las oclusivas del español pronunciadas por aprendices sinohablantes de dos perfiles lingüísticos y comparar los valores con la pronunciación de hispanohablantes.

En este estudio han participado 11 sujetos que se pueden clasificar en tres grupos. El grupo 1 está formado por cinco alumnas nativas de chino mandarín y el grupo 3 está constituido por dos estudiantes nativas del dialecto wu del chino. En el momento de la grabación cursaban estudios de posgrado en español en la Universidad Autónoma de Barcelona, lo cual garantiza su buen dominio de la lengua española. El grupo 2 está formado por cuatro hispanohablantes que son alumnas universitarias y proceden de la provincia de Barcelona. El corpus de grabación, similar a los que utilizan Castañeda (1986) y Rosner *et al.* (2000), está formado por 150 palabras bisílabas aisladas que contienen las oclusivas en cuestión. Las oclusivas solamente aparecen en ataque silábico y en una de las dos posiciones: en sílaba tónica (como /b/ en *vaca*) o en sílaba átona (como /b/ en *vagar*). A fin de controlar la categoría gramatical, todas las palabras son de

[3] El término de VOT (*voice onset time* en inglés, Lisker y Abramson 1964) se refiere al tiempo que transcurre entre la apertura de la oclusión de una oclusiva y el comienzo de la vibración de las cuerdas vocales (Hualde 2005, XXI).

contenido (sustantivo, verbo y adjetivo). Para evitar que errores de pronunciación se produzcan a causa del desconocimiento, en la confección del corpus se ha tenido en cuenta la frecuencia de aparición basándose en la estadística proporcionada por el banco de datos NIM (Guasch *et al.* 2013), un buscador que facilita estímulos experimentales para la investigación en psicolingüística.

La grabación se realizó en una cabina insonorizada en dicha universidad. A cada informante se le proporcionó una serie de diapositivas en PowerPoint, que se exponían en una pantalla situada a 1,5 metros del informante. En cada diapositiva se exhibía una sola palabra y el informante debía leerla con voz natural. La diapositiva cambiaba cada 3 segundos; de esta forma, se pretendía que los individuos pronunciasen al estilo de citación y con una velocidad regulada. Para la grabación, se utilizó un micrófono cardioide de condensador AudioTechnica AT2020, situado a 15-20 cm. de la boca del informante. Una vez obtenido el archivo original, se segmentó el audio y se midieron los parámetros acústicos manualmente a través del programa Praat. Concretamente, el VOT se midió en el oscilograma desde el inicio de sonoridad de baja amplitud y frecuencia hasta la explosión en el caso de las sonoras y desde el inicio de la primera onda aperiódica que indica la explosión hasta el comienzo del primer formante vocálico en el caso de las sordas.

## 3 Resultados

### 3.1 Comparación del valor de VOT dentro de cada grupo

Las cinco alumnas del grupo 1 pronunciaron 326 casos de /p t k/ y 251 casos de /b d g/. Todas las /p t k/ tienen valor positivo y 100 de las /b d g/ tienen valor negativo, lo cual supone un 39,8 % del total. En la Tabla 1 se exponen los valores correspondientes a las oclusivas que tienen VOT positivo. Como se puede observar, el VOT suele ser más largo cuando la oclusiva se encuentra en sílaba átona.

| Oclusiva | /p/ | | /t/ | | /k/ | |
|---|---|---|---|---|---|---|
| Sílaba | tónica | átona | tónica | átona | tónica | átona |
| Media | 9.68 | 10.82 | 12.96 | 13.71 | 24.32 | 24.03 |
| SD | 4.14 | 5.30 | 4.98 | 5.23 | 8.63 | 9.65 |
| Oclusiva | /b/ | | /d/ | | /g/ | |
| Sílaba | tónica | átona | tónica | Sílaba | tónica | átona |
| Media | 9.18 | 9.27 | 13.34 | Media | 9.18 | 9.27 |
| SD | 4.23 | 4.07 | 4.24 | SD | 4.23 | 4.07 |

Tabla 1. VOT de las oclusivas del grupo 1 con VOT positivo

En la Tabla 2 se exhiben los valores de /b d g/ que tienen sonoridad anticipada. Es necesario señalar que 82 de estos 100 casos corresponden a los individuos 1 y 2, mientras que los demás pronuncian /b d g/ con sonoridad anticipada esporádicamente. La duración del VOT negativo de oclusivas incrustadas en sílabas átonas es menor que en sílabas tónicas.

| Oclusiva | /b/ | | /d/ | | /g/ | |
|---|---|---|---|---|---|---|
| Sílaba | tónica | átona | tónica | átona | tónica | átona |
| Media | -97.87 | -88.47 | -121.37 | -98.92 | -101.34 | -84.68 |
| SD | 49.81 | 51.71 | 37.43 | 44.98 | 26.30 | 40.53 |

Tabla 2. VOT de las oclusivas del grupo 1 con VOT negativo

Según el resultado de la prueba Kruskal-Wallis, el VOT medio de /k/ es significativamente superior al de /p/ (H=160,83, p<0.001) y al de /t/ (H=108,62, p<0,001). Adicionalmente, el VOT medio de /t/ resulta significativamente superior al de /p/ (H=52,21, p<0,001). De manera similar, en /b d g/ con VOT positivo se dan estos contrastes: el VOT medio de /k/ es significativamente superior al de /p/ (H=84,73, p<0,001) y al de /t/ (H=49,78, p<0,001), que, a su vez, es superior al de /p/ (H=34,94, p<0,001). En los casos de VOT negativo, la diferencia que reside entre las tres series no resulta significativa (H=1,66, p=0,434). Asimismo, la variable de posición no implica diferencia significativa en ninguna serie de oclusivas a pesar de las diferencias observadas.

Dado que /p t k/ y /b d g/ sin sonoridad anticipada tienen valores de VOT similares, se realizaron una serie de pruebas Mann-Whitney para examinar si existe diferencia relevante entre cada par de oclusivas homoorgánicas /p-b/, /t-d/ y /k-g/. Los resultados señalan que en las bilabiales (U=3400, p=0,08) y las alveolares (U=2218, p=0,312) no se dan diferencian significativas; y en las velares el VOT medio de /g/ es incluso significativamente más largo que el de /k/ (U=829, p<0,001). Estos resultados apuntan inequívocamente a la confusión de categoría en los individuos 3, 4 y 5.

Los cuatro integrantes del grupo 2 pronuncian 556 oclusivas iniciales en total. Los 323 casos de /p t k/ tienen valor positivo de VOT y 231 casos de /b d g/ tienen sonoridad anticipada. Dos pronunciaciones de /d/ poseen VOT positivo, por lo que se consideran como casos esporádicos.

| Oclusiva | /p/ | | /t/ | | /k/ | |
|---|---|---|---|---|---|---|
| Sílaba | tónica | átona | tónica | átona | tónica | átona |
| Media | 14.27 | 15.81 | 15.75 | 19.51 | 28.78 | 33.29 |
| SD | 6.30 | 8.01 | 5.97 | 7.56 | 8.20 | 9.03 |
| Oclusiva | /b/ | | /d/ | | /g/ | |
| Sílaba | tónica | átona | tónica | átona | tónica | átona |
| Media | -82.39 | -68.64 | -83.24 | -69.47 | -71.74 | -62.78 |
| SD | 21.50 | 17.37 | 24.77 | 21.21 | 24.60 | 20.90 |

Tabla 3. VOT de las oclusivas del grupo 2

En las oclusivas sordas, el resultado de pruebas ANOVA indica que el VOT medio de las velares es significativamente superior al de las bilabiales (F=15,37, p<0,001) y al de las alveolares (F=12,75, p<0,001). Adicionalmente, el VOT medio de las alveolares es más largo que el de las bilabiales (F=2,62, p=0,035). Respecto a la variable del acento, es obvio que el VOT medio es más largo cuando la oclusiva está en sílaba átona, aunque esta diferencia carece de significación (t= -0,87, p=0,321). En cambio, en las oclusivas sonoras no se encuentran diferencias relevantes en el VOT en las tres series (F=1,492, p=0,227). En lo referente al acento, se observa que la sonoridad anticipada es más larga cuando la oclusiva está en sílaba tónica. Esta diferencia es significativa en /b/ (t= -3,90, p=0,039), pero no en /d/ (t= -2,76, p=0,321) ni en /g/ (t= -1,23, p=0,5).

Finalmente, los dos informantes del grupo 3 pronuncian un total de 212 casos de oclusivas iniciales, de los cuales 87 corresponden a /b d g/. Solamente 3 casos de /b/, un caso de /d/, un caso de /p/ y un caso de /t/ tienen sonoridad anticipada, y todas las demás oclusivas se pronuncian con VOT positivo. Estos casos no se incluyeron en el análisis estadístico, ya que pueden distorsionar la distribución de los valores del total.

| Oclusiva | /p/ | | /t/ | | /k/ | |
|---|---|---|---|---|---|---|
| Sílaba | tónica | átona | tónica | átona | tónica | átona |
| Media | 10.20 | 10.66 | 10.23 | 10.67 | 20.78 | 19.70 |
| SD | 3.49 | 4.38 | 2.91 | 2.21 | 7.53 | 4.80 |
| Oclusiva | /b/ | | /d/ | | /g/ | |
| Sílaba | tónica | átona | tónica | átona | tónica | átona |
| Media | 8.06 | 7.24 | 10.37 | 11.37 | 23.94 | 31.85 |
| SD | 3.28 | 2.26 | 3.02 | 2.66 | 11.30 | 7.45 |

Tabla 4. VOT de las oclusivas del grupo 3

Según las pruebas Kruskal-Wallis, en /p t k/ pronunciadas por los dos informantes del grupo 3, /k/ tienen un VOT medio significativamente más largo que /p/ ($H=7,5$, $p<0,001$) y /t/ ($t=6,32$, $p<0,001$). Si bien en /p/ y /k/ la posición postónica se asocia a un VOT más largo, en ninguna serie la diferencia implicada por la posición resulta significativa. En /b d g/ que tienen VOT positivo, las mismas pruebas corroboran la significatividad de la diferencia en VOT implicada por las distintas series. El VOT medio de las velares es significativamente superior al de /b/ ($H=48,20$, $p<0,001$) y al de /d/ ($H=27,35$, $p=0,001$) que, a su vez, es significativamente más largo que el de las bilabiales ($H=20,85$, $p=0,002$). De manera similar, la posición tampoco conlleva diferencia relevante.

Al igual que en el grupo 1, se realizaron pruebas Mann-Whitney para comparar los valores de VOT de /p t k/ con los de /b d g/ que, mayoritariamente, se pronuncian sin sonoridad anticipada. Los resultados indican que mientras el VOT medio de /b/ es significativamente inferior al de su contraparte /p/ ($U=467$, $p<0,001$), el de /g/ es superior al de /k/ ($U=424$, $p=0,037$), mientras que entre /d/ y /t/ no existen diferencias significativas.

## 3.2 Comparación del valor de VOT entre los grupos

Una vez aclarados los valores correspondientes a cada segmento oclusivo en cada grupo, la comparación entre los tres grupos resulta necesaria. Anteriormente, se ha señalado que dos informantes del grupo 1 pronuncian la mayoría de /b d g/ con sonoridad anticipada, mientras los demás lo hacen ocasionalmente. Por ello, es necesario separar los datos correspondientes a estos dos informantes (al que denominaremos grupo 1.1) del resto (grupo 1.2).

| | G3 | G2 | G1.2 |
|---|---|---|---|
| G1.1 | /b/**[4] /d/** /g/** /k/* | /p/** /t/** /k/* | /b/** /d/** /g/** /t/** |
| G1.2 | /t/** | /b/** /d/** /g/** /p**/ /k/** | |
| G2 | /b/** /d/** /g/** /p/** /t/** /k/** | | |

Tabla 5. Resumen de significación de las comparaciones en VOT entre los grupos

[4] Los dos asteriscos ** indican que el nivel de significación es inferior a 0,001, mientras que un asterisco * indica que el nivel de significación es inferior a 0,05.

La Tabla 5 resume la significación de las comparaciones correspondiente a cada oclusiva entre los cuatro grupos. Se observa que entre el grupo 2 y el 3 la diferencia es significativa en todas las oclusivas; entre el grupo 2 y el 1.2, hay diferencias significativas en todas las oclusivas menos /t/; sin embargo, no se obtienen diferencias relevantes entre el grupo 2 y el 1.1 en las sonoras /b d g/. El grupo 1.1 difiere del 1.2 y del 3 en todas las oclusivas sonoras y determinadas sordas (/t k/), mientras que entre el grupo 1.2 y el 3 la única diferencia significativa reside en /t/. Asimismo, es conveniente señalar que la pronunciación de las oclusivas de los dos individuos del grupo 1.1 es la más cercana a la nativa en términos de VOT, diferenciándose así de la pronunciación de los demás aprendices sinohablantes. La confusión de las oclusivas sonoras con las sordas sigue constituyendo la mayor dificultad para estos informantes, aunque los valores de VOT de las sordas tampoco coinciden con los nativos. Esta observación también puede verse en la Figura 1 que recoge los valores medios concernientes a cada oclusiva y a cada grupo.

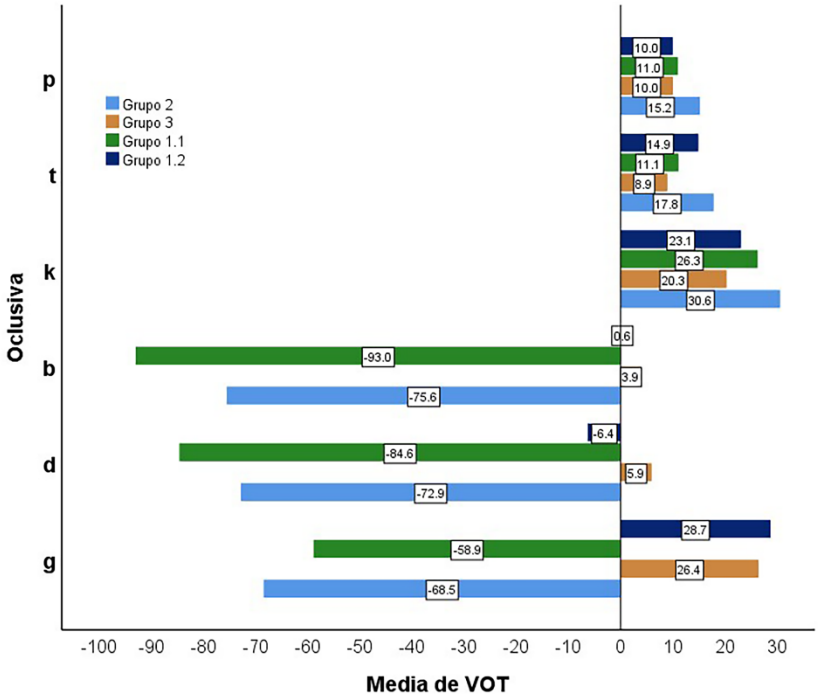


Figura 1. Valores de VOT según oclusiva y grupo

## 4 Discusión

A continuación, se discuten los resultados obtenidos enfocando los siguientes aspectos: en primer lugar, se comparan los valores absolutos con otros estudios; en segundo lugar, se pretende explicar los resultados según el marco teórico; finalmente, se intenta analizar la notable diferencia individual observada.

### 4.1 Comparación con otros estudios

Según los datos recogidos en la Tabla 6, los valores de /p t k/ del grupo de hispanohablantes (grupo 2) son superiores a los de Castañeda (1986) y a los de Rosner *et al.* (2000), mientras que los valores de los grupos sinohablantes (grupos 1 y 3) son

más cercanos a los de Castañeda. Respecto a las oclusivas sonoras /b d g/, los datos del grupo 2 son próximos a los de Castañeda y superiores a los de Rosner *et al.*, mientras que los del grupo 1.1 son similares a los de Rosner *et al.,* excepto en el caso de /g/.

| | /p/ | /t/ | /k/ | /b/ | /d/ | /g/ |
|---|---|---|---|---|---|---|
| Castañeda (1986) | 6,5 | 10,4 | 25,7 | -69,8 | -77,7 | -58 |
| Rosner *et al.* (2000) | 13,1 | 14,0 | 26,5 | -91,5 | -91,6 | -73,7 |
| Grupo 2 | 14,53 | 17,41 | 30,06 | -75,57 | -72,86 | -68,48 |
| Grupo 1.1 | 8,27 | 12,18 | 27,51 | -94,04 | -84,64 | -58,9 |
| Grupo 1.2 | 8,08 | 12,38 | 24,09 | 0,63 | -6,35 | 28,71 |
| Grupo 3 | 10,46 | 10,45 | 20,31 | 7,71 | 10,84 | 26,41 |

Tabla 6. VOT de los grupos del presente estudio y de otros estudios citados

A pesar de que los valores no sean exactamente similares, en los grupos 2 y 1.1 las tendencias que destacan coinciden efectivamente con los citados estudios. Por un lado, el VOT de las sordas aumenta a medida que retrocede el lugar de articulación y la sonoridad anticipada de /g/ es la más corta. Por otro lado, la diferencia en el VOT entre las oclusivas sordas y sus contrapartes sonoras resulta estable dentro de cada grupo de hablantes, tal y como se observa en la Tabla 7. Los grupos 1.2 y 3 sí presentan tendencias similares en las sordas y la incapacidad de producción de /b d g/ con sonoridad anticipada impide la existencia de una diferencia eficaz entre las sordas y las sonoras, tal y como se ha visto con anterioridad.

| | /p-b/ | /t-d/ | /k-g/ |
|---|---|---|---|
| Castañeda (1986) | 76,3 | 88,1 | 83,7 |
| Rosner *et al.* (2000) | 104,6 | 105,6 | 100,2 |
| Grupo 2 | 90,1 | 90,27 | 98,54 |
| Grupo 1.1 | 102,31 | 96,82 | 86,41 |

Tabla 7. VOT de los grupos de este estudio y de Castañeda (1986) y Rosner *et al.* (2000)

## 4.2 Mecanismos de aprendizaje según *Speech Learning Model*

Los aprendices del grupo 1.2 y del 3 cometen errores de confusión de categoría de manera similar y coinciden en mostrar una incapacidad total de producir /b d g/ con sonoridad anticipada. Si bien esta dificultad del grupo 1 se puede atribuir a la inexistencia de sonidos oclusivos sonoros, el contraste fonológico de sonoridad existente en la L1 de los individuos del grupo 3 no parece contribuir a la adquisición de las oclusivas sonoras castellanas. Además, es necesario reconocer la notable diferencia individual que reside en los individuos del grupo 1 que comparten la misma L1, así como una experiencia de aprendizaje similar.

Tanto el éxito del grupo 1.1 como la dificultad de los demás informantes pueden explicarse según el *Speech Learning Model*. De acuerdo con dicho modelo (Flege 1995, 239), las dos alumnas del grupo 1.1 distinguen exitosamente las oclusivas sordas de las sonoras porque han logrado separar perceptivamente /b d g/ del castellano de /p t k/ nativas, que son sonidos más parecidos a aquellos. Este mecanismo es imprescindible para la formación de una nueva categoría de sonidos en una L2/LE. De modo contrario, los errores del grupo 1.2 se producen a causa de que estos aprendices clasifican perceptivamente las sonoras castellanas como “similares” a /p t k/ en su L1.

Parece extraño que las oclusivas sordas no aspiradas /p t k/, cuyo VOT es ligeramente superior a cero, y las sonoras con sonoridad anticipada /b d g/ se perciban como parecidas. En realidad, en la mayoría de las lenguas habladas en el país asiático, como el mandarín y el cantonés, la sonoridad anticipada no constituye fonológicamente un rasgo que tiene capacidad distinguidora. En este sentido, aunque /b d g/ castellanas y /p t k/ del chino mandarín difieran significativamente en su VOT, a un nativo del chino mandarín le cuesta percibir que son dos categorías de sonidos distintas. Este tipo de confusión de categoría también puede verse en la pronunciación de las oclusivas sonoras castellanas (Flege y Eefting 1988) por estadounidenses, quienes apenas pronuncian las sonoras del español puertorriqueño adecuadamente en tareas de imitación. La confusión se debe a que, al igual que en el chino mandarín, las oclusivas del inglés tampoco se distinguen por su sonoridad anticipada en posición inicial absoluta (Ladefoged y Johnson 2011, 151).

Respecto a la confusión observada en el grupo 3, el hecho de que el contraste fonológico de sonoridad en las oclusivas de su L1 no favorezca el aprendizaje de /b d g/ españolas pone de manifiesto que la transferencia de la L1 a la L2/LE se produce en un nivel superficial en lugar de en un nivel fonémico. Si bien los segmentos sonoros /b d g/ son existentes en el dialecto wu, su manifestación acústica difiere considerablemente de las /b d g/ españolas. Según se ha estudiado (Chen y Gussenhoven 2015, 324), el VOT de una oclusiva sonora del dialecto wu no difiere significativamente de las sordas no aspiradas /p t k/ cuando se encuentra al inicio de una unidad prosódica, y el contraste fonológico se manifiesta fonéticamente en el modo de fonación y en la F0 de la vocal adyacente, y no en la misma consonante. En líneas del modelo *Speech Learning Model* (Flege 1995, 239), los sonidos de la L1 y de la L2 se relacionan perceptualmente en un nivel alofónico que es sensible a la posición, en vez de en un nivel fonémico más abstracto. Asimismo, las /b d g/ castellanas iniciales se asocian perceptivamente, bien como /b d g/, bien como /p t k/ de la L1 de los informantes del grupo 3; como resultado de esta asociación, la sonoridad anticipada de las oclusivas de la L2/LE pasa desapercibida, lo cual constituye la causa de la incapacidad de producir las /b d g/ castellanas de manera adecuada.

### 4.3 Diferencia individual en el aprendizaje

Según se mencionó anteriormente, la diferencia entre la pronunciación de los sujetos del grupo 1.1 y del 1.2 pone de manifiesto la relevante diferencia individual dentro de un conjunto de alumnos de naturaleza homogénea en cuanto a aspectos como su L1, edad, tiempo de aprendizaje, nivel de formación, entre otros. Dicha diferencia no debe atribuirse a factores como velocidad de elocución, variación diastrática y desconocimiento de las palabras, ya que todas las informantes leen el mismo corpus en las mismas condiciones experimentales.

A este respecto, hay que admitir que es difícil rastrear la causa de este fenómeno sobre la base del resultado de este estudio. Por un lado, el éxito del aprendizaje depende en cierta manera de una compleja serie de capacidades personales y factores extralingüísticos, como el estilo de aprendizaje, la estrategia y la motivación. Por otro lado, el hecho de que se desconozca cómo perciben los sonidos en cuestión rechaza la posibilidad de estudiar si el error de pronunciación se produce a causa de que la percepción de estos sonidos sea defectuosa. Una razonable hipótesis sobre la diferencia individual podría ser que los dos sujetos del grupo 1.1 hayan recibido mayor cantidad de *input* audiovisual, como la radio y la televisión, suficiente como para que su capacidad

distinguidora haya mejorado perceptivamente y que, en consecuencia, resulte en una producción más adecuada.

## 5 Conclusión

Este estudio constituye un intento de describir y contrastar la pronunciación de las oclusivas castellanas iniciales por el alumnado sinohablante de dos perfiles lingüísticos diferentes. Los resultados señalan que ciertos individuos nativos del chino mandarín son capaces de producir estos sonidos de manera más satisfactoria, pero los demás sujetos de esta L1 y los dos sujetos nativos del dialecto wu cometen errores de confusión de manera similar. No obstante, la causa de la confusión de los nativos del chino mandarín es la inexistencia de segmentos oclusivos sonoros en su L1, mientras la diferencia en la manifestación fonética del mismo contraste fonológico entre el español y el wu es responsable de la dificultad de aprendizaje de los aprendices de este perfil lingüístico. Además, la relevante diferencia individual observada recuerda la necesidad de reconocer la existencia de diversidad dentro de un alumnado que parece homogéneo y de atender los factores individuales de aprendizaje para garantizar un buen resultado.

### REFERENCIAS BIBLIOGRÁFICAS


Castañeda Vicente, M.L. (1986). El V.O.T. de las oclusivas sordas y sonoras españolas. *Estudios de Fonética Experimental*, 2, 92-110.

Chen, Y. y Gussenhoven, C. (2015). Shanghai Chinese. *Journal of the International Phonetic Association*, 45(3), 321-227. doi:10.1017/S0025100315000043.

Flege, J.E. (1995). Second language speech learning: Theory, findings and problems. En W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 233–277). Timonium: York Press.

Flege, J. E. y Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America*, 83(2), 729–740. https://doi.org/10.1121/1.396115.

Gil, J. (2007). *Fonética para profesores de español: de la teoría a la práctica*. Madrid: Arco Libros.

Guasch, M., Boada, R., Ferré, P. y Sánchez-Casas, R. (2013). NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods*, 45, 765-771.

Hualde, J.I. (2005). *The sounds of Spanish*. Nueva York: Cambridge University Press.

Ladefoged, P. y Johnson, K. (2011). *A course in phonetics*. Wadsworth: Cengage Learning.

Lee, W. y Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33:1, 109-112. doi:10.1017/S0025100303001208.

Lee, W. y Zee, E. (2014). Chinese Phonetics. En C.T.J. Huang, Y.H.A. Li y A. Simpson (Eds.), *The handbook of Chinese linguistics* (pp. 369–399). Nueva Jersey: Wiley Blackwell.

Lisker, L. y Abramson, A.S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *WORD*, 20(3), 384-422. https://doi.org/10.1080/00437956.1964.11659830.

Martínez Celdrán, E. y Fernández Planas, A.M. (2007). *Manual de fonética española*. Barcelona: Ariel.

Quilis, A. (1981). *Fonética acústica de la lengua española*. Madrid: Gredos.

Rosner, B.S., López-Bascuas, L.E., García-Albea, J.E. y Fahey, R.P. (2000). Voice-onset times for Castilian Spanish initial stops. *Journal of Phonetics*, 28, 217-224. https://doi.org/10.1006/jpho.2000.0113.