

## La percepción auditiva en la interfaz entre Fonética, Fonología y ELE. Las oclusivas del español en aprendices sinohablantes

*Auditory perception at the Phonetics-Phonology-SFL interface.  
Spanish stops by Chinese-speaking learners*

ALEJANDRA PADILLA-BROM

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO / UNIVERSIDAD NACIONAL DE  
EDUCACIÓN A DISTANCIA  
apadilla@cepe.unam.mx / apadilla142@alumno.uned.es

VICTORIA MARRERO-AGUIAR

UNIVERSIDAD NACIONAL DE EDUCACIÓN A DISTANCIA  
vmarrero@flog.uned.es

### Resumen:

En este trabajo se presentan los resultados de un estudio sobre la discriminación auditiva de los contrastes entre consonantes oclusivas orales del español por aprendices sinohablantes, analizando los datos mediante la teoría de detección de señales. Tras unos apartados iniciales de contextualización, tanto teórica como metodológica, y de un breve análisis contrastivo entre las lenguas implicadas, se presentan los resultados de una prueba de discriminación de respuesta cerrada, utilizando el paradigma igual/diferente, que confirman las dificultades de estos aprendientes para distinguir la serie sorda o tensa de la sonora o laxa. Por ello se ha diseñado una propuesta didáctica específica destinada a mejorar este aspecto crítico en la pronunciación del español como lengua extranjera.

**Palabras clave:** percepción auditiva, adquisición de ELE, oclusivas del español, aprendices sinohablantes, teoría de detección de señales

### Abstract:

This paper presents the results of a study on the auditory discrimination of oral Spanish consonant-occlusive contrasts by Sino-speaking learners, analysing the data using signal detection theory. After some initial theoretical and methodological contextualisation sections and a brief contrastive analysis between the languages involved, the results of a closed-response discrimination test, using the same/different paradigm, are presented,



ISSN: 2462-3733

n.º 11 | Diciembre 2025 | 156-173  
DOI: <https://doi.org/10.5565/rev/doblele.177>  
Universitat Autònoma de Barcelona

Recibido: 29/9/2025  
Aprobado: 1/12/2025

confirming the difficulties of these learners in distinguishing the voiceless or tense series from the voiced or voiceless or lax series. Therefore, a specific didactic proposal has been designed to improve this critical aspect in the pronunciation of Spanish as a foreign language.

**Keywords:** auditory perception, Spanish L2 acquisition, Stop consonants, Chinese-speaking learners, Signal Detection Theory

## 1. Introducción

El colectivo de aprendices de ELE sinohablantes es uno de los que más ha crecido en las últimas décadas, tanto en China (Álvarez, 2020; Man, 2021; Zhai & Álvarez, 2020) como en distintos países hispanohablantes, entre ellos, España o México, y tanto en niveles universitarios como en los de escolarización primaria y secundaria. Por lo tanto, no es de extrañar el creciente interés de los profesores por mejorar sus estrategias y propuestas didácticas para este grupo.

Entre las principales dificultades de los aprendices chinos están, evidentemente, las gramaticales; pero la pronunciación también les plantea retos importantes, que a menudo genera problemas de fosilización, porque es, frecuentemente, la habilidad lingüística que menor atención recibe en las aulas.

Uno de esos escollos habituales de este colectivo es nuestro objeto de estudio: los contrastes de oclusivas orales /p t k b d g/. En este trabajo, que forma parte de otro más amplio, nos limitaremos solo a la discriminación auditiva de las realizaciones obstruyentes. Presentaremos, en primer lugar, los resultados de un estudio de discriminación auditiva de sílabas, y en la segunda parte del trabajo, una propuesta didáctica para mejorar su percepción.

## 2. Marco teórico

La percepción auditiva de los sonidos del habla en el contexto del aprendizaje de español como lengua extranjera (ELE) es un área de investigación en la que se integran los hallazgos de la fonética perceptiva, la fonología y la adquisición de segundas lenguas. Las dos últimas son disciplinas tradicionales que no requieren definición; la fonética perceptiva, de desarrollo más reciente, es la disciplina lingüística encargada de estudiar los mecanismos de interpretación de los sonidos del habla (audición, percepción y comprensión) en relación con el conjunto del sistema lingüístico (Marrero-Aguiar, 2021). Esos mecanismos son fundamentales para comprender cómo los aprendices de segundas lenguas procesan y discriminan los contrastes fonológicos del español y su lengua materna (L1), ya que ayudan a explicar cómo interactúan ambos sistemas fonológicos.

Diversas teorías han intentado dar cuenta de esos procesos. El modelo *Automatic Selective Perception* (ASP) (Strange, 2011; Strange & Shafer, 2008) sostiene que los adultos perciben la L1 de forma automatizada gracias a las rutinas de percepción selectiva (SPRs, por sus siglas en inglés) que han desarrollado a lo largo de su vida; con ellas identifican los sonidos y contrastes fonéticos de la L1, incluso en situaciones de escucha complejas o cuando no están poniendo atención. En cambio, afirma que los “aprendientes tardíos” no perciben la L2 de forma automática, es decir, les representa un esfuerzo cognitivo mayor, pues necesitan escuchar de forma

consciente para identificar los contrastes fonéticos que no existen en su lengua.

La interferencia de la L1 aparece cuando los aprendices aplican de forma automática las rutinas perceptivas de su lengua materna en el momento de procesar los sonidos de la L2, lo que dificulta distinguir los contrastes fonológicos inexistentes en su L1, generando confusiones o errores de percepción. Para lograr un aprendizaje exitoso de la L2 es necesario favorecer nuevas rutinas perceptivas mediante un entrenamiento intenso enfocado a la discriminación de sonidos, ya que la simple exposición natural a la L2 no garantiza su desarrollo. También el *Speech Learning Model (SLM)* (Flege, 1995) y su versión revisada (*SLM-r*) (Flege & Bohn, 2021) plantean que el sistema fonético de la L1 influye en la percepción de los sonidos de la L2, ya que estos tienden a ser percibidos mediante el filtro de las categorías fonéticas existentes en la L1. Pero los aprendientes tendrían la capacidad de formar nuevas categorías fonéticas para sonidos de la L2, siempre y cuando perciban claramente las diferencias entre los sonidos: a mayor diferencia, mayor probabilidad de crear una categoría nueva. Estos modelos apuntan que el factor edad puede dificultar este proceso: cuanto más tarde se comience el aprendizaje de la L2, menor será la capacidad para aprender los nuevos sonidos. Así pues, proponen que existen diferencias individuales en la percepción tanto de la L1 como de la L2 y que están relacionadas con la precisión de las categorías fonológicas, la experiencia auditiva y la cantidad y calidad de la exposición a la L2. Estos modelos sostienen que la percepción y la producción de los sonidos de la L2 evolucionan juntas y se influyen mutuamente, sin que una preceda necesariamente a la otra.

Por su parte, el *Perceptual Assimilation Model (PAM)* (Best, 1995; Best et al., 2001) explica que la L1 moldea la percepción de la L2 a través de la producción; en otras palabras, la experiencia con la L1 empuja a los oyentes a asimilar la L2 a sus categorías fonológicas nativas, en un proceso basado en los gestos motores necesarios para articular esos sonidos. Los sonidos no nativos se interpretan en relación con su similitud o diferencia con los patrones gestuales de los sonidos de la L1. El *Second Language Linguistic Perception Model (L2LP)* (Escudero, 2005; van Leussen & Escudero, 2015), no sólo confirma la importancia de los factores perceptivos ya mencionados, sino que incorpora elementos adicionales como el procesamiento léxico y el aprendizaje basado en el significado. Este modelo explica que cuando se aprende una L2, los individuos perciben la lengua nueva mediante el filtro de su L1 y, a medida que el aprendizaje avanza, crean nuevas categorías perceptuales para los sonidos inexistentes en su L1. Para lograrlo, el *L2LP* propone un entrenamiento fonético contrastivo por medio del cual se destaque las diferencias entre L1 y L2. Además, sostiene que el aprendizaje puede optimizarse al utilizar el léxico como guía, pues cuando el vocabulario se amplía, el reconocimiento de las palabras impulsa ajustes finos en la gramática perceptiva del estudiante, con ello, gradualmente perfecciona su capacidad de escuchar.

Finalmente, mencionaremos también las aportaciones del método verbotonal (Carrera-Sabaté et al., 2021; Guberina, 1961; Padilla, 2018; Renard, 1979) aunque se trata de una propuesta más metodológica que teórica: predominio de la percepción, la prosodia y la gestualidad en etapas iniciales, evitando interferencias de la escritura; atención al componente afectivo; y, en el nivel segmental, control de la relación entre la tensión de los sonidos y su posición silábica y de los efectos coarticulatorios.

En suma, nuestro estudio de la percepción de consonantes oclusivas (/p, b, t, d, k, g/) en aprendices sinohablantes de español se enmarca en la línea de los modelos que destacan el papel de la percepción en el desarrollo de la competencia fonológica en segundas lenguas.

### 3. Marco metodológico. La teoría de detección de señales (TDS)

Desde el punto de vista metodológico, en este trabajo utilizamos una teoría psicofísica clásica perfectamente asentada en otros ámbitos de la ciencia, pero menos difundida en lingüística<sup>1</sup>. Actualmente se utiliza de forma habitual en campos de conocimiento como la medicina o la psicología, la teoría de detección de señales (TDS) (Green & Swets, 1966; Hautus et al., 2021; Wickens, 2002).

Para explicar la TDS es necesario considerar los cuatro tipos de respuesta posibles ante una tarea de detección. Aunque hay diferentes paradigmas posibles, utilizaremos como ejemplo el utilizado en este trabajo: igual/diferente (*same/different*). Los participantes indican si los dos miembros de una pareja de estímulos (en nuestro caso, sílabas con una consonante oclusiva y una vocal) son iguales (por ejemplo, dos emisiones de [te]), o si son distintos (una es [te] y otra, [de]). Las respuestas se clasifican en cuatro tipos:

*Tabla 1. Clasificación de las respuestas en tareas de detección con el paradigma igual/diferente*

	Estímulos iguales	Estímulos distintos
Respuesta "iguales"	Rechazo correcto	Error
Respuesta "distintos"	Falsa alarma	Acierto

A partir de aquí, los puntos de partida de la TDS son los siguientes:

- Cualquier presentación de estímulos puede provocar un cambio en la respuesta de los sujetos, pero siempre existen distorsiones que introducen variabilidad en la señal (fatiga, cambios de atención, estado físico o fisiológico, ruido...). Y eso explica que en ambas condiciones (estímulos iguales y distintos) podamos

<sup>1</sup> A pesar de que, curiosamente, Swets & Birdsall aplicaron la TDS al reconocimiento perceptivo en el mismo simposio, celebrado en 1956 en el MIT, en el que Chomsky planteó las bases de la gramática generativa.

encontrar las dos respuestas (“son distintos” o “son iguales”); eso es algo que depende de la señal, pero también de otros muchos factores circunstanciales. En definitiva, no hay efectos sensoriales constantes o estables.

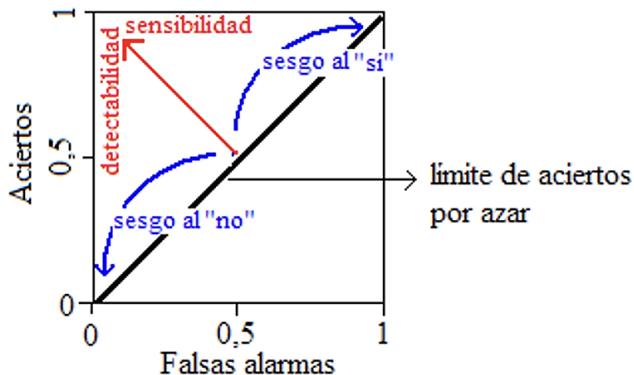
- Todo proceso de detección de señales comporta dos subprocesos, uno de origen sensorial, dependiente de la sensibilidad del sujeto y la detectabilidad de la señal, y otro de origen cognitivo, dependiente de su criterio de respuesta (o sesgo): las personas más arriesgadas tenderán a dar más respuestas “son distintos” y las más conservadoras darán más respuestas “son iguales”. Así pues, el receptor se considera un elemento activo en el proceso de detección de las señales, y sus mecanismos “de arriba abajo” inciden de forma muy importante en las respuestas.
- Se pueden separar los efectos originados por cada uno de los dos subprocesos. Esta es la gran aportación de la TDS, la posibilidad de diferenciar la sensibilidad del sujeto (indicada mediante el signo  $d'$ , “d prima”) de su tendencia a responder “sí” o “no”, de su carácter más o menos conservador, del nivel de certeza que necesita (“c”, criterio de respuesta, o  $\beta$ , una variante del mismo). Para conseguirlo, basta considerar la proporción entre aciertos y falsas alarmas (Tabla 1). La medida de sensibilidad es la distancia entre la media de dos distribuciones: la de las falsas alarmas y la de los aciertos<sup>2</sup>.

Esos datos además se pueden representar gráficamente en las llamadas curvas ROC (*receiver operative curve*, curva operativa de respuesta o curvas de isosensibilidad). En el eje horizontal situamos la proporción de falsas alarmas; en el eje vertical, la de aciertos. La línea diagonal de la gráfica separa los aciertos por azar de las respuestas motivadas; si el experimento se ha desarrollado correctamente (y el oyente ha contestado de forma coherente), no habrá respuestas por debajo de esa diagonal. Cuanto más a la izquierda encontraremos las tasas de respuesta, mayor será la tendencia del sujeto a contestar “no” ante la presentación de los estímulos, más conservadoras serán sus respuestas; por el contrario, hacia la derecha tenderán los sujetos más arriesgados, los que, en caso de duda, optarían con frecuencia por el “sí”. Por último, cuanto más arriba aparezca la respuesta, mayor será la sensibilidad del sujeto o la detectabilidad de la señal.

---

<sup>2</sup>  $d'$  es la diferencia entre la media de las dos distribuciones (la de aciertos y la de las falsas alarmas) dividida entre la desviación típica común a esas dos distribuciones.

Figura 1. Curva operativa de respuesta  
(receiver operative curve, ROC)



Una vez planteados los marcos teórico y metodológico de este trabajo, concluiremos estos apartados iniciales con un resumen contrastivo de los subsistemas fonológicos que conocen nuestros aprendices: las consonantes oclusivas de la lengua meta, el español, y las de la lengua materna, el chino, sin perder de vista las del inglés, su primera lengua extranjera.

#### 4. Descripción contrastiva de las consonantes oclusivas españolas, chinas e inglesas

El español cuenta con dos series de fonemas oclusivos orales: /p t k/ y /b d g/. Cada una de estas series presenta, respectivamente, puntos de articulación bilabial, dental y velar. /p t k/ se realizan como oclusivas sordas sin aspiración, tanto entre vocales como en inicio de palabra (Martínez & Padilla, 2021: 44); por su parte, /b d g/ tienen dos alófonos en distribución complementaria: se realizan oclusivas después de una pausa, de una consonante nasal y, en el caso de /d/, después de una lateral (/l/); en el resto de los contextos se presentan como aproximantes [β, ð, γ], principalmente entre vocales y, para /d/, después de otras consonantes que no sean nasales ni laterales (Hualde, 2013). La diferencia entre /p t k/ y /b d g/ se atribuye generalmente al rasgo sordo/sonoro (cuyas claves acústicas son tanto la barra de sonoridad como el VOT, *voice onset time*), pero, además /p t k/ son sonidos más tensos que /b d g/ (Machuca, 1995; Martínez & Fernández, 2007; Quilis, 1993; Veiga & Arias, 2024).

El chino mandarín cuenta también con seis fonemas oclusivos orales distribuidos igualmente en dos series, pero todos ellos son sordos o tensos, la diferencia entre ambas series está en la aspiración: /p t k/ frente a /p<sup>h</sup> t<sup>h</sup> k<sup>h</sup>/ (Cortés, 2009; Liu, 1991; Yufei, 2007). Los alófonos aproximantes tampoco aparecen en el inventario fonológico del mandarín, aunque sí en variedades o lenguas sínicas como el wu, de Shanghai (Yufei, 2007).

Por otra parte, la mayoría de los aprendices sinohablantes de español tienen conocimientos previos del sistema fonológico del inglés, donde encontramos la oposición fonológica entre sordas o fuertes /p t k/ y sonoras

o laxas /b d g/ (sin alófonos aproximantes), pero fonéticamente, en las sílabas tónicas las sordas se articulan con aspiración ([p<sup>h</sup> t<sup>h</sup> k<sup>h</sup>]) (Giegerich, 1992; Pennington, 2014; Roach, 2009).

Todo ello hace que esta serie de fonemas constituya uno de los retos más señalados en los aprendices sinohablantes (Chen, 2007; Chih, 2014; Cortés, 2014; Gong et al., 2017; Wei, 2025).

## 5. Estudio empírico: prueba perceptiva

### 5.1. Metodología

#### *Participantes*

Los participantes son 18 sinohablantes, 10 hombres y 8 mujeres. Sus edades van de los 18 a los 43 años. Nueve tienen estudios universitarios, la misma cantidad terminó la preparatoria. Diez son monolingües; tres son bilingües (mandarín-inglés) y cinco son trilingües (dos, en mandarín —“chino nativo”— e inglés; uno, en mandarín, inglés y eslovaco; otro, en mandarín, inglés y japonés; y uno en mandarín y otras dos lenguas sínicas). Todos ellos se encuentran realizando cursos de español, en el nivel inicial (A2 del MCER) en el servicio de idiomas de la UNAM.

#### *Tarea y estímulos. Discriminación de consonantes sordas y sonoras*

El propósito de esta tarea es evaluar si nuestros aprendientes pueden distinguir auditivamente entre sílabas con consonantes oclusivas del español, tanto sordas como sonoras (/p b t d k g/). Para ello, se aplicó una prueba de percepción auditiva que incluyó 24 ítems, 12 correspondían a pares de sílabas diferentes, es decir, con contraste fonológico (por ejemplo, /pe/-/be/, /ke/-/ge/), y otros 12 presentaban la misma sílaba en dos emisiones diferentes (por ejemplo, /pe/-/pe/, /ke/-/ke/). Los estímulos se presentaron aleatorizados como sigue:

*Tabla 2. Descripción de los estímulos*

N.º	Estímulo	N.º	Estímulo	N.º	Estímulo	N.º	Estímulo
1	[ke] [ge]	7	[de] [de]	13	[te] [de]	19	[pe] [pe]
2	[be] [be]	8	[pe] [be]	14	[de] [de]	20	[pe] [pe]
3	[te] [te]	9	[ke] [ge]	15	[te] [de]	21	[te] [de]
4	[pe] [be]	10	[ke] [ge]	16	[be] [be]	22	[ge] [ge]
5	[te] [de]	11	[te] [te]	17	[pe] [be]	23	[ke] [ge]
6	[ke] [ke]	12	[ge] [ge]	18	[ke] [ke]	24	[pe] [be]

La instrucción y el audio de ejemplo eran estos:

Figura 2. Instrucción y ejemplo de la tarea

A continuación, vas a escuchar pares de sílabas. Indica "iguales" si son la misma sílaba o "diferentes" si son dos sílabas diferentes. Puedes escuchar varias veces cada audio.

Ejemplos:

Respuesta: diferentes

Respuesta: iguales

接下来您将听到一些音节对。如果听到的音节相同，请回答"相同"; 如果不同，请回答"不同"。您可以重复收听每段音频。

示例：

答案：不同

答案：相同

Las respuestas se registraron en formato dicotómico (1 = acierto, 0 = error).

## 5.2. Resultados

### 5.2.1. Resultados generales

La tasa de respuestas correctas de la tarea fue de 64,3 %, con un promedio de 15 sobre un máximo de 24 posibles.

Tabla 3. Resultados generales

	%	Promedio	Desviación estándar (DE)
Respuestas acertadas	278	64,35	15,44
Respuestas erróneas	154	35,65	8,56

Para determinar si la diferencia entre aciertos y respuestas no acertadas fue significativa, una vez comprobada la distribución normal de la muestra (prueba de Shapiro-Wilk,  $p$ -valor = 0.9092) aplicamos una prueba  $t$  de Student para muestras no emparejadas (puesto que los participantes realizaron la tarea de forma individual), comparando la tasa de error/acierto de los 18 participantes. El resultado arrojó un valor  $p < 0.005$ , indicando que la tasa de aciertos fue significativamente más alta que la de errores.

A continuación, desglosamos los datos diferenciando entre pares contrastivos y pares iguales (en cuyo caso los aciertos se denominan rechazos correctos y los errores, falsas alarmas). Obtuvimos los siguientes resultados:

Tabla 4. Resultados por categorías

	Pares contrastivos	Pares iguales	
Aciertos	146 (67,59 %)	Rechazos correctos	132 (61,11 %)
Errores	70 (32,41 %)	Falsas alarmas	84 (38,89 %)

Las diferencias entre los resultados de pares contrastivos y pares iguales fueron muy pequeñas; al aplicar de nuevo la prueba *t de Student*, en este caso para muestras emparejadas (porque cada estudiante resolvió ambos tipos de ítems); el resultado, como era de esperar, fue que las diferencias no son significativas ( $p = 0.467$ ). Por lo tanto, las tasas de errores (pares contrastivos identificados como iguales) y falsas alarmas (pares iguales identificados como contrastivos) son estadísticamente iguales.

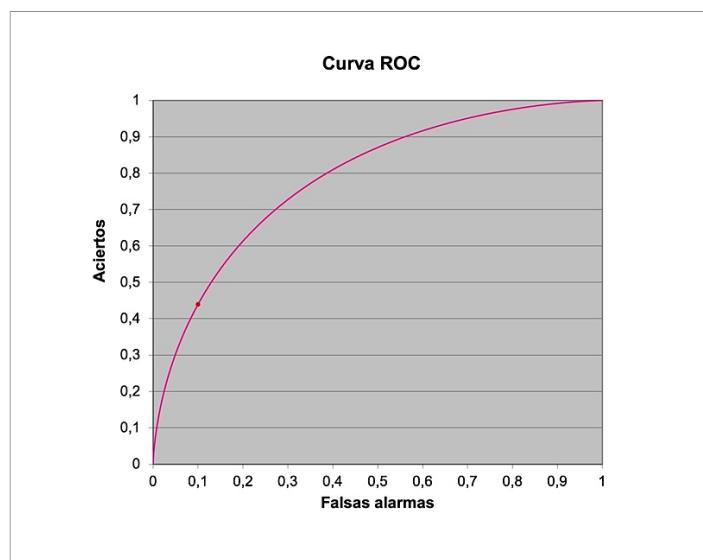
Dado el diseño de la prueba, se aplicó la TDS para obtener la curva ROC y los valores de sensibilidad, criterio y sesgo de respuesta<sup>3</sup>:

*Tabla 5. Valores de sensibilidad y criterio*

d'	0,786
$\beta$	0,938

Un valor  $d' = 0,74$  supone una capacidad baja para discriminar los contrastes analizados, aunque superior al acierto por azar. Un valor de  $\beta$  cercano a 1 indica que no ha habido sesgos en las respuestas de los participantes, solo una leve tendencia a contestar “diferentes” aunque los estímulos fueran iguales.

*Figura 3. Curva ROC de la tarea de discriminación de consonantes sordas y sonoras*



<sup>3</sup> Estos cálculos y la curva ROC de la Figura 3 se han obtenido con la aplicación del Dr. Moreno Pedraza. Disponible en línea: <https://www.ucm.es/serranopedrazalab/software-1> [23/09/2025].

Por lo tanto, nuestros resultados confirman que los contrastes entre consonantes oclusivas sordas y sonoras del español suponen un reto para los aprendices de la muestra analizada, aunque su tasa de acierto, posiblemente como resultado de las clases de pronunciación recibidas, está significativamente por encima de la de errores, y del acierto por azar.

### **5.2.2. /p t k/ versus /b d g/**

La tasa de aciertos de la serie sorda y la sonora son casi idénticos (65 y 60 respectivamente, sobre un total de 108). Considerando solo las repeticiones de estímulos iguales (porque en los pares distintivos siempre había un elemento sordo y otro sonoro, de modo que no podemos atribuir el acierto o el error a uno solo de los elementos de la pareja), las falsas alarmas se distribuyeron de forma prácticamente equilibrada entre la serie sorda, /p t k/ (43 falsas alarmas, sobre un total de 108 ítems), y la sonora /b d g/ (41 falsas alarmas).

### **5.2.3. Lugares de articulación**

Si desglosamos las respuestas considerando las diferencias entre consonantes bilabiales (/p b/), dentales (/t d/) y velares (/k g), estos son los resultados:

*Tabla 6. Resultados por lugares de articulación*

	Bilabiales <i>/pe/-/be/</i>	Dentales <i>/te/-/de/</i>	Velares <i>/ke/-/ge/</i>
Total resp. acertadas	97 (67,4 %)	97 (67,4 %)	84 (58,3 %)
Total resp. erradas	47 (32,6 %)	47 (32,6 %)	60 (41,7 %)
Aciertos	48 (66,7 %)	58 (80,6 %)	40 (55,6 %)
Errores	24 (33,3 %)	14 (19,4 %)	32 (44,4 %)
Rechazos correctos	49 (68,1 %)	39 (54,2 %)	44 (61,1 %)
Falsas alarmas	23 (31,9 %)	33 (45,8 %)	28 (38,9 %)
d'	0,90	0,97	0,17
β	1,02	0,69	0,94

La estadística descriptiva apunta a que las consonantes velares resultaron más difíciles para nuestros aprendices: una diferencia de un 10 % entre el total de respuestas acertadas de los pares /k/-/g/, y las de dentales y labiales (que, curiosamente, coinciden exactamente en su porcentaje) no parece muy grande, pero la prueba chi-cuadrado sobre los datos binarios indican que esa diferencia es significativa ( $p < 0.005$ ). La TDS nos confirma que, efectivamente, el valor de  $d'$  en esta categoría es muy inferior al que presentan labiales y dentales, que están en el límite de una sensibilidad moderada ( $d'$  próxima a 1).

La diferencia entre labiales y dentales aparece al desglosar esos datos globales en las cuatro categorías de análisis de la TDS: en las primeras

la tasa de respuestas acertadas fue la misma tanto en los pares contrastivos como en los iguales (48 % de aciertos, 49 % de rechazos correctos); sin embargo, en las dentales se dispara la respuesta “distintos”, tanto cuando los estímulos efectivamente lo son (58 % de aciertos) como cuando en realidad son iguales (33 % de falsas alarmas, la tasa más alta de todas); por eso el sesgo de respuesta de las dentales es claramente inferior a 1, indicativo de una tendencia “liberal” respecto al “sí, hay diferencia”, frente a labiales y velares, cuyo valor de “c” ronda el 1 (indicativo de que no hay sesgo de respuesta).

### 5.3. Conclusiones de la prueba perceptiva

Los resultados obtenidos en una prueba perceptiva de discriminación de sílabas indican que la capacidad de nuestro grupo de aprendices sinohablantes para diferenciar las consonantes oclusivas sordas y sus correlatos sonoros está por encima del acierto por azar, pero bastante lejos de una discriminación óptima, o incluso moderada: un 64 % de respuestas acertadas, que suponen un promedio de 15, sobre un máximo de 24. Por eso su sensibilidad ( $d'$ ) no alcanza el 0,8 (en esta tarea el valor máximo, correspondiente a una sensibilidad altísima, sería superior a 5).

La responsable principal de esa baja discriminabilidad en los contrastes de las consonantes oclusivas orales es la serie velar: la diferencia entre [ke] y [ge] apenas supera el 50 % de acierto por azar. Además, en la serie dental aparece un sesgo en las respuestas, una tendencia a contestar “diferente” tanto cuando la pareja que se ha de discriminar es [te]-[de], como cuando es [de]-[de] o [te]-[te]. Curiosamente, el contraste labial, que se suele considerar el error prototípico en sinohablantes (y que resultó la serie peor percibida en el trabajo de Liu, 2019: 24 % de error, frente al 11 % de la velar y el 3 % de la dental) fue, en nuestra muestra, la que obtuvo mejores resultados. Las diferencias con estudios previos pueden deberse a las características de los participantes (nuestro grupo ha recibido instrucción formal sobre pronunciación) y de la prueba (en Liu, 2019 se utilizaron palabras bisílabas y una tarea de identificación), aunque no puede descartarse la influencia del tamaño de la muestra.

## 6. Propuesta didáctica

A continuación, presentamos una propuesta didáctica para el nivel A2 del MCER, cuyo objetivo es proporcionar un entrenamiento intensivo de discriminación (Escudero, 2005; Strange, 2011) que facilite también mejoras en la producción de los contrastes entrenados: que la percepción auditiva sirva como base para una comprensión y producción oral significativas. El tema transversal es la organización de una fiesta, una actividad integrada en la programación de aula general. En su aplicación a nuestra muestra se incidirá particularmente en los contrastes entre oclusivas velares, /k g/, que han resultado los más difíciles para nuestros aprendices.

## 6.1. Actividad rompe hielo

- Objetivo: activar conocimientos previos, crear un ambiente relajado y preparar la discriminación auditiva.
- Descripción: se presentan grabaciones de sonidos relacionados con una fiesta y su organización, e imágenes que representan la fuente del sonido.
- Materiales: audios de música, risas, brindis de copas, el timbre de una puerta, gente conversando animadamente, explosión de un globo, el descorche de una botella, etc., e imágenes asociadas.
- Tarea: emparejamiento de sonido e imagen.

## 6.2. Escucha con propósito específico

- Objetivo: activar el conocimiento lingüístico e indexical.
- Descripción: se presentan secuencias de palabras y frases de diferentes locutores (distintas edades, hombres y mujeres, monólogos y conversaciones grupales) con una fiesta, para que se identifiquen las lenguas, características individuales del locutor y del contexto.
- Material: audios cortos con palabras y oraciones como: "Vino, comida, dulces" / "¿Quieres una copa?"; *wine, food, sweets / Are you coming to the party?*; *Pútáojiǔ, shíwù, tángguǒ / Nǐ yào lái cānjiā jùhuì ma?*
- Tarea: seleccionar la respuesta correcta en un cuestionario de respuesta cerrada con opciones como las siguientes:

Tabla 7. Hoja de respuestas del ejercicio 6.2.

¿En qué lengua habla?	¿Quién habla?
	
Chino 	
Ingles 	
Español 	
Otro 	

### 6.3. Sensibilización articulatoria

- Objetivo: reconocer las propiedades articulatorias de las consonantes oclusivas /p b t d k g/.
- Descripción: el profesor produce estos sonidos mediante gestos articulatorios exagerados, para hacer patentes los lugares de articulación y la diferencia de sonoridad; se utilizan espejos y velas para la retroalimentación visual/táctil, cuidando de no generar una aspiración en las oclusivas del español que sí aparecer en las del chino.
- Material: imágenes de esquemas articulatorios, como los que aparecen en la página web *Sounds of speech*, de la Universidad de Iowa<sup>4</sup>, personalizados para indicar la vibración o no de las cuerdas vocales<sup>5</sup>; espejos pequeños individuales y velas.
- Tarea: imitación corporeizada ante el espejo (dedos de la mano izquierda en la garganta, y de la derecha en los labios para /p b/, dientes para /t d/ y articulación mandibular para /k g/).

### 6.4. Sensibilización acústico-auditiva

- Objetivo: reconocer las propiedades acústicas y auditivas de las consonantes oclusivas /p b t d k g/.
- Descripción: se presentan las consonantes oclusivas del español, bien como sonidos aislados (ruido explosivo con o sin sonoridad), bien en sílabas con la vocal /e/; se puede utilizar el paradigma AXB empleado en el experimento perceptivo anterior, pidiendo que se identifiquen los dos sonidos iguales, o bien que se detecte el diferente (*odd one out*). Inicialmente, solo se escucha el sonido. Luego, se añade labiolectura (por parte del profesor o mediante un video) y finalmente, se muestra la grafía (en color rojo, las correspondientes a sonidos sonoros y en azul, las de los sordos).
- Material: grabaciones de los seis fonemas y un cuestionario de tipo test.
- Tarea: detección auditiva mediante cuestionario de respuesta cerrada.

### 6.5. Contrastes de pares mínimos

- Objetivo: identificar unidades léxicas que solo se diferencian en un fonema oclusivo (sordo o sonoro), en sílaba inicial o final.
- Descripción: en primer lugar, se presentan las palabras con las que se va a trabajar con su correspondiente significado, a través

<sup>4</sup> Disponible en línea: <https://soundsofspeech.uiowa.edu/spanish> [23/09/2025].

<sup>5</sup> Los sonidos sonoros se marcan con una onda roja vibrante, y los sordos con una ráfaga de aire azul.

de imágenes representativas. A continuación, se emparejan las imágenes y se explica la tarea. Cuando los aprendices están preparados, se presentan las grabaciones de los pares mínimos aleatorizados.

- Material: grabaciones de los siguientes pares mínimos en los dos órdenes posibles; imágenes asociadas, presentadas en el contexto de una fiesta:

*peso - beso; paso - vaso; pala - bala; pela - vela; pata - bata  
tenso - denso; toma - doma; tuna - duna; teja - deja; tardo - dardo  
coma - goma; cata - gata; casa - gasa; cala - gala; quise - guise*

Para incrementar la dificultad de la tarea se pueden utilizar pares mínimos donde el contraste se dé, en lugar de en la sílaba inicial tónica, en la final átona, que es menos tensa y perceptible:

*capa - cava; lapa - lava; ropa - roba; cupo - cubo  
bota - boda; mota - moda; cata - cada; coto - codo  
toca - toga; saca - saga; loco - logo; placa - plaga*

- Tarea: hay diferentes posibilidades, de dificultad variable. La más compleja sería la de identificación y repetición: elegir la imagen correspondiente al estímulo escuchado y repetirlo; una versión más fácil sería crear tríos de estímulos (dos iguales y uno diferente) para una tarea de detección de intruso (*odd one out*), que puede combinar labiolectura y presentación de grafías. También se puede pedir, simplemente, clasificar las tarjetas en dos columnas: sonidos sordos (asociados al icono azul del soplo) y sonidos sonoros (icono rojo de la onda sonora).

## 6.6. Oclusivas en contexto

- Objetivo: generalizar los aprendizajes de las tareas estructuradas anteriores a contextos más libres.
- Descripción: se presentan, bien mediante grabación, bien de viva voz, frases cortas relacionadas con el tema de la fiesta, en la que aparecen algunas de las palabras trabajadas anteriormente.
- Material: frases como las siguientes:

“Toma el vaso”, “pon la vela en la tarta”, “lavó la copa de cava”, “saca a la gata de casa, que no se coma la goma”, etc.

- Tarea: en la versión más fácil se incluye en la frase una palabra repetida que se debe identificar: “toma el vaso, te paso ese vaso”, “pon la vela en la tarta, la vela de boda”, “saca a la gata de casa, de cada casa”, etc. Una tarea de dificultad intermedia

consistiría en responder mediante mímica con la acción indicada (el gesto de coger un vaso, el de colocar una vela, de lavar una copa, etc.). La más compleja sería repetir la frase y grabarla, para escuchar y valorar posteriormente la pronunciación.

## 7. Conclusiones generales

Este trabajo tiene un doble objetivo: didáctico y metodológico. Por lo que respecta al primero, se ha partido de un análisis de necesidades (detectado mediante un test de discriminación, que ha puesto de manifiesto las dificultades de nuestro grupo meta para discriminar los contrastes entre oclusivas orales sordas y sonoras) para elaborar una propuesta didáctica basada en el refuerzo de la percepción, que responde a los principios de modelos teóricos como el *Automatic Selective Perception* (ASP) (Strange, 2011; Strange & Shafer, 2008) o el *Second Language Linguistic Perception Model* (Escudero, 2005; van Leussen & Escudero, 2015). Tras un par de actividades de acercamiento afectivo y finalidad atencional, se parte de ejercicios de sensibilización, acústico-auditiva y articulatoria (como defiende el *Perceptual Assimilation Model*, Best, 1995 y Best *et al.*, 2001); a continuación se utiliza un léxico muy controlado, en pares mínimos, para afianzar la identificación del contraste estudiado, que finalmente se generaliza a contextos más amplios, tanto de percepción como de producción.

En cuanto al objetivo metodológico, la tarea de detección (paradigma igual/diferente) que hemos utilizado nos ha permitido aplicar a los resultados las propuestas de la TDS (Green & Swets, 1966; Hautus *et al.*, 2021; Wickens, 2002), y determinar así la sensibilidad de nuestros aprendices a los contrastes analizados. Hemos llegado a la conclusión de que las oclusivas orales del español les plantean un reto importante, porque su capacidad general para discriminárlas es baja, y particularmente el contraste entre consonantes velares, /k/- /g/. La TDS se creó para explicar cómo se detectan señales en presencia de ruido, entendido no solo en sentido literal, sino también entendido subjetivamente, como incertidumbre. De sus orígenes en la ingeniería y la psicofísica (Green & Swets, 1966) ha pasado a aplicarse de forma generalizada en medicina (McFall & Treat, 1999), neurociencia (Wixted, 2020), e incluso en inteligencia artificial (Langer *et al.*, 2023). Este trabajo es una pequeña muestra de su utilidad para la lingüística aplicada.

## 8. BIBLIOGRAFÍA

- BEST, C. T. (1995): "A direct realist view of cross-language speech perception", en W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues*, pp. 171-204. York Press.
- BEST, C. T., MCROBERTS, G. W., & GOODELL, E. (2001): "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's

- native phonological system". *The Journal of the Acoustical Society of America*, 109(2), pp. 775-794.
- CARRERA-SABATÉ, J., LÓPEZ GARCÍA, P., & BACH-MARQUÈS, J. (2021): *Nous horitzons del mètode verbotal: Millorar la pronúncia holísticament*. Graó.
- CHEN, Y. (2007): *A comparison of Spanish produced by Chinese L2 learners and native speakers. An acoustic phonetics approach* [Tesis doctoral]. University of Illinois at Urbana-Champaign.
- CHIH, M. T. (2014): "Errores de percepción fonética en estudiantes taiwaneses de ELE". *Revista SinoELE. Suplementos 10*.
- CORTÉS MORENO, M. (2009). *Fonología china*. Herder.
- CORTÉS MORENO, M. (2014): "Dificultades lingüísticas del español para los estudiantes sinohablantes y búsqueda de soluciones motivadoras", en *La enseñanza del español para sinohablantes en contextos*, pp. 173-208. Monográficos SinoELE.
- ESCUDERO, P. (2005): *Linguistic perception and second language acquisition* [Tesis doctoral]. Utrecht University.
- FLEGE, J. E. (1995): "Second language speech learning: Theory, findings, and problems". *Speech perception and linguistic experience: Issues in cross-language research*, 92, pp. 233-277.
- FLEGE, J. E., & BOHN, O. S. (2021): "The Revised Speech Learning Model (SLM-r)", en R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress*, pp. 3-83. Cambridge University Press.
- GIEGERICH, H. J. (1992): *English phonology: An introduction*. Cambridge University Press.
- GONG, J., LECUMBERRI, M. L. G., & COOKE, M. (2017): "Ab initio perceptual learning of foreign language sounds: Spanish consonant acquisition by Chinese learners". *System*, 66, pp. 142-155.
- GREEN, D. M., & SWETS, J. A. (1966): *Signal detection theory and psychophysics*. Wiley.
- GUBERINA, P. (1961): "La méthode audio-visuelle structuro-globale et ses implications dans l'enseignement de la phonétique". *Studia Romanica et Anglica Zagabiensia*, 11, pp. 12-40.
- HAUTUS, M. J., MACMILLAN, N. A., & CREELMAN, C. D. (2021): *Detection theory: A user's guide*. Routledge.
- HUALDE, J. I. (2013): "Consonantes oclusivas", en J. I. Hualde, *Los sonidos del español. Spanish Language edition*, pp. 129-145. Cambridge University Press.
- LANGER, M., BAUM, K., & SCHLICKER, N. (2023, 31 julio): "A signal detection perspective on error and unfairness detection as a critical aspect of human oversight of AI-based systems". PsyArXiv, <https://doi.org/10.31234/osf.io/ke256> (consultado el 25/09/25)
- LIU, N. (2012): *Análisis de errores y evolución de la interlengua de los estudiantes sinohablantes: esbozo de un estudio longitudinal* [Tesis doctoral]. Universidad Autónoma de Barcelona.
- LIU, Z. (2019): *Análisis de las obstruyentes en chino y en español como L3: Estudio acústico y perceptivo para la categorización de errores* [Tesis doctoral]. Universidad Autónoma de Barcelona.
- MCFALL, R. M., & TREAT, T. A. (1999): "Quantifying the information value of clinical assessments with signal detection theory". *Annual review of psychology*, 50(1), pp. 215-241.
- MACHUCA AYUSO, M.ª J. (1997): *Las obstruyentes no continuas del español: relación entre las categorías fonéticas y fonológicas en el habla espontánea* [Tesis doctoral]. Universitat Autònoma de Barcelona.

- MAN, Y. (2021): El español en China. En *El español en el mundo: Anuario del Instituto Cervantes*, pp. 657-683. Instituto Cervantes.
- MARRERO-AGUIAR, V. (2021): “La perspectiva perceptiva. El lado sumergido en el iceberg de la fonética” [Seminario científico]. *Doctorado en Español: Investigación Avanzada en Lengua y Literatura*, 28 de abril. Universidad de Salamanca.
- MARTÍNEZ-BELDA, J. C., & PADILLA, X. A. (2021): “VOT in Spanish stops in simulated emotional speech”. *Journal of Experimental Phonetics*, 30, pp. 35-57. <https://doi.org/10.1344/efe-2021-30-35-57>
- MARTÍNEZ CELDRÁN, E., & FERNÁNDEZ PLANAS., A. M.<sup>a</sup> (2007): *Manual de fonética española*. Ariel.
- MILLER, G. A. (2006): “La revolución cognitiva: una perspectiva histórica”. *Revista de Psicología*, 25(2), pp. 79-88.
- PADILLA GARCÍA, X. (2018): “Remarks on verbo-tonal phonetics for a communicative context”. *Normas*, 8, pp. 259-271. <https://doi.org/10.7203/Normas.v8i1.13433>
- PENNINGTON, M. C. (2014): *Phonology in English language teaching: An international approach*. Routledge.
- QUILIS, A. (1993): *Tratado de fonología y fonética españolas*. Gredos.
- RENARD, R. (1979): *Introduction à la méthode verbo-tonale de correction phonétique*. CIPA.
- ROACH, P. (2009): *English Phonetics and Phonology: A Practical Course*. Cambridge University Press.
- STRANGE, W. (2011): “Automatic selective perception (ASP) of first and second language speech: A working model”. *Journal of Phonetics*, 39(4), pp. 456-466. <https://doi.org/10.1016/j.wocn.2010.09.001>
- STRANGE, W. Y SHAFER, V. (2008): “Speech perception in second language learners: The re-education of selective perception”, en J. G. Hansen Edwars & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition*, pp. 153-191. John Benjamins.
- VAN LEUSSEN, J. W., & ESCUDERO, P. (2015): “Learning to perceive and recognize a second language: The L2LP model revised”. *Frontiers in psychology*, 6, Article 1000. <https://doi.org/10.3389/fpsyg.2015.01000>
- VEIGA RODRÍGUEZ, A., & ARIAS CABAL, A. (2024): “Descripción fonológica de las consonantes oclusivas orales”, en J. Gil Fernández & J. Llisterri, *Fonética y fonología descriptivas de la lengua española*, pp. 412-443. Georgetown University Press.
- ZHAI, H., & ÁLVAREZ ÁLVAREZ, M. (2020): “El Español como Lengua Extranjera (ELE) con el fin específico del comercio exterior en la Educación Vocacional y Aplicada en China: una revisión sistemática del estado de la cuestión”. *Méjico y la Cuenca del Pacífico*, 9(26), pp. 117-137.
- WEI, Y. (2025): “Stop production of L1 Chinese - L2 English - L3 Spanish learners: The source and direction of crosslinguistic influence”. *Journal of Experimental Phonetics*, 34, pp. 21-40. <https://doi.org/10.1344/efe-2025-34-21-40>
- WICKENS, T. D. (2002): *Elementary signal detection theory*. Oxford University Press.
- WIXTED, J. T. (2020): “The forgotten history of signal detection theory”. *Journal of experimental psychology: learning, memory, and cognition*, 46(2), pp. 201-233. <https://doi.org/10.1037/xlm0000732>
- YUFEI, C. (2007): “Un estudio contrastivo de los fonemas oclusivos entre español y chino”. *Méjico y la Cuenca del Pacífico*, 10(28), pp. 91-98. <https://doi.org/10.32870/mycp.v10i28.289>