LA SELECCIÓN DE JUECES NATIVOS PARA LA EVALUACIÓN DEL ACENTO EXTRANJERO EN APRENDICES FRANCÓFONOS DE ESPAÑOL L2

MARGARITA PLANELLES ALMEIDA ANNA DOQUIN DE SAINT PREUX
Universidad Antonio de Nebrija

RESUMEN

La existencia de aprendices adultos de segundas lenguas que alcancen un nivel de dominio de la L2 similar al nativo es una cuestión muy controvertida en el campo de la adquisición de segundas lenguas, tradicionalmente ligada a las discusiones en torno a la existencia de un periodo crítico para la adquisición de segundas lenguas. Esta cuestión es especialmente polémica en lo que se refiere a la adquisición de la competencia fonológica, esto es, a la existencia de aprendices con un nivel de producción oral en la L2 "libre de acento extranjero". La utilización de jueces nativos para la evaluación del acento extranjero en hablantes de una L2 es un procedimiento extendido pero no exento de polémica. En los últimos años, diferentes autores han señalado la necesidad de establecer criterios comunes y fundamentados para la evaluación de la competencia fonológica en L2. El presente estudio se propone explorar las diferencias en la evaluación del acento extranjero entre jueces nativos no expertos (naïve) y jueces nativos experimentados. Los participantes fueron 20 jueces nativos, clasificados en dos grupos en función de su nivel de conciencia respecto a la percepción del acento extranjero (12 jueces inexpertos, y 12 jueces experimentados en la enseñanza del español como lengua extranjera). Por otro lado, se seleccionó un grupo de 20 aprendices francófonos de español con diferentes edades de inicio de adquisición, todos ellos con un tiempo mínimo de 7 años de residencia en España y un alto nivel de pronunciación, así como un grupo de 6 nativos que actuó como grupo de control. Se evaluó el grado de acento extranjero en la pronunciación de los informantes a través de los dos paneles de jueces nativos. Para el análisis de los resultados, en primer lugar, se estableció la fiabilidad de los jueces a la hora de reconocer el habla nativa, excluyendo aquellos jueces que fueron incapaces de valorar como tales al grupo de

nativos introducido como grupo de control. A continuación, se realizaron pruebas de validez intra-jueces (correlación de las puntuaciones otorgadas por el mismo juez a los mismos informantes en diferentes momentos). De nuevo, aquellos jueces que no resultaron fiables fueron excluidos del estudio. Por último, se compararon los resultados obtenidos por los informantes en la evaluación por uno y otro panel de jueces (inexperto vs. experimentado).

Palabras clave: Acento extranjero; Competencia fonológica/ Pronunciación; Jueces nativos naïves o inexpertos; Jueces nativos experimentados; Español L2

1. INTRODUCCIÓN

La cuestión del llamado "acento extranjero", como uno de los rasgos más identificativos de la producción oral de los hablantes de una L2, ha sido ampliamente discutida en la literatura sobre la adquisición de segundas lenguas. La posibilidad de que un aprendiz tardío (de una L2) pueda realmente alcanzar un nivel de dominio de la L2 a nivel nativo es una de las cuestiones más controvertidas en este campo. Especialmente, el llamado "acento extranjero" parece ser el reducto ineliminable incluso para aquellos aprendices que presentan un nivel de dominio excepcional, comparable al nativo (*nativelike*), en otros aspectos de la lengua. Tradicionalmente ligados al debate sobre la existencia de un periodo crítico o sensible, son numerosos los estudios que evalúan el grado de acento extranjero en las producciones orales de diferentes aprendices de una L2, tratando de testar esta hipótesis. Para ello, numerosos estudios utilizan paneles de jueces nativos para cuantificar el grado de acento extranjero de los informantes. Sin embargo, la utilización de estos jueces para la evaluación del acento extranjero en hablantes de una L2 es un procedimiento, si bien enormemente extendido, no por ello exento de polémica. En los últimos años, diferentes voces han señalado, justamente, la necesidad de ahondar en las cuestiones metodológicas que subyacen a este tipo de investigaciones, y de llegar a algún tipo de consenso sobre cómo establecer "whether two populations – (e.g., monolingual natives and L2 learners) perform "the same" or "differently" in what is arguably the most challenging task in (late) L2 learning, namely attaining a native accent in L2 pronunciation." (Schmid y Hopp, 2014, p. 368). El presente trabajo se centra en una de estas cuestiones concernientes a las decisiones

metodológicas a la base de estos estudios: la utilización de jueces lingüísticamente *naïve*s o *inexpertos* (sin entrenamiento en lingüística o fonética, sin familiaridad con las lenguas maternas de los aprendices), *versus* la evaluación por jueces *expertos* (con formación específica y familiaridad con el aprendizaje de lenguas extranjeras).

2. HACIA UN CONSENSO METODOLÓGICO

Son numerosos los estudios científicos que, desde el planteamiento inicial de la existencia de un periodo crítico o sensible para la adquisición de segundas lenguas (Lenneberg, 1967), han tratado de validar o refutar esta hipótesis atendiendo a los diferentes niveles o dominios de la lengua (morfosintaxis, fonética/fonología, vocabulario, pragmática). Concretamente, numerosos autores señalan la evidente dificultad para alcanzar una pronunciación libre de acento nativo cuando el aprendizaje de la L2 comienza a una edad tardía: "the critical period hypothesis (CPH) is widely viewed as providing an explanation for why many individuals speak their L2 with a foreign accent." (Flege, 1999, p. 102).

Diferentes estudios presentan, en este sentido, evidencias a favor de la existencia de tal limitación intrínseca (Asher y García, 1969; Oyama, 1976; Suter, 1976; Seliger, Krashen y Ladefoged, 1975; Tahta, Wood y Loewenthal, 1981a; Purcell y Suter, 1980; Patkowski, 1990; Thompson, 1991; Flege, 1988; Flege, Yeni-Komshian y Lu, 1999; Hyltenstam y Abrahamsson 2003b; Abrahamsson y Hyltenstam, 2009; Abrahamsson, 2012; Granena y Long, 2013). Sin embargo, no son pocos los autores que, por su parte, presentan casos de aprendices tardíos con un nivel de pronunciación comparable al nativo (Bongaerts, Planken y Schills, 1997; Flege, Munro y MacKay, 1995; Bongaerts, 1999; Bongaerts, Mennem y van der Slik, 2000; Birdsong, 2003; Moyer, 1999; Granfeldt-Saavedra, 2007; Birdsong, 2007; Colantoni y Steele, 2006; Nagle, 2013).

En efecto, la investigación (muy numerosa) sobre el tema presenta resultados muy diferentes, y, aunque la existencia de aprendices tardíos con una pronunciación nativa constituye una

contra-evidencia frente a la hipótesis del periodo crítico, la interpretación de los resultados existentes, así como las conclusiones a las que estas llevan, son muy distintas -a menudo, radicalmente opuestas- en unos autores y en otros. En este sentido, algunos consideran la proliferación de casos de aprendices excepcionales como el resultado de errores o falta de criterios estrictos en la metodología y los procedimientos de toma de datos y evaluación utilizados (Abrahamsson y Hyltenstam, 2009; Long, 2005) En relación con esto, diferentes investigadores han subrayado la importancia de llegar a algún tipo de consenso metodológico, esto es, de fijar o estandarizar en cierto modo los procedimientos metodológicos de recolección de muestras, operativización de las variables e instrumentación, evaluación de los resultados…; una estandarización metodológica que permita, en definitiva, una verdadera comparación de los resultados y sirva así para poder avanzar en la verdadera comprensión de los fenómenos subyacentes, pues, como defiende DeKeyser (2013, p. 59), "…it is important to assess the L2 learners' knowledge in appropriate ways to elicit the kind of knowledge that matters to the theory". En este sentido, el autor apunta a la necesidad de establecer criterios metodológicos (de selección, evaluación y análisis de datos) comunes y realiza algunas sugerencias para su mejora, "in the hope that both sides of the debate will be able to agree on them." (DeKeyser, 2013, p. 52) Pues, en efecto, a pesar de la importancia central del debate en torno a la posibilidad de alcanzar un nivel de dominio nativo en la L2 y de sus implicaciones tanto teóricas como prácticas, "the exact nature of the "native target" as well as questions of how to establish, measure and assess it, have received relatively little attention" (Schmidd y Hopp, 2014, p. 368).

Respecto a la competencia fonológica existen, como explican Schmid y Hopp (2014), dos tipos de enfoques fundamentales para evaluar hasta qué punto un aprendiz de L2 ha alcanzado un nivel nativo (un habla "libre de acento extranjero"):

The first applies measurements either to the output [...] or to articulatory settings during speech production [...] The second consists

of eliciting native speaker ratings of segments of speech produced by L2 learners for their accentedness. (p. 368)

Ambos métodos presentan, por supuesto, ventajas e inconvenientes (ver Schmid y Hopp, 2014). En lo que aquí nos interesa, la evaluación del grado de acento extranjero (global) por paneles de jueces es, sin lugar a dudas, un procedimiento ampliamente utilizado en las investigaciones existentes. No obstante, la variabilidad en las decisiones metodológicas es, en ellas, cuanto menos llamativa, cuando no confusa y, a menudo, ni tan siquiera explícitamente presentada y/o justificada (para una revisión sobre el tema, ver Jesney, 2004, y Schmid y Hopp, 2014). Así, como destaca Jesney (2004):

> Foreign accent, as an identifying aspect of nonnative speech, has been widely discussed in both the theoretical and pedagogical literature. Still, the elements that contribute to the perception of foreign accent, and, indeed, the objective characteristics of it, remain ill defined. With this in mind, numerous studies have sought to quantify degrees of foreign accentedness by asking groups of listeners to rate speech samples for degrees of foreign accent using a Likert scale. […] The precise methodology utilized in making such foreign accentedness ratings remains unstandardized, however, and, as a consequence, numerous differences in both specific approaches and findings have been noted across studies. (p. 2)

Una mirada rápida a la bibliografía revela, en efecto, la gran variabilidad existente en este tipo de estudios en cuestiones que podrían ser determinantes, a saber: el número y características de los informantes (el número general de los informantes, sus lenguas maternas, edad, y la información proporcionada sobre sus *backgrounds*); las muestras de habla (que van desde la lectura en voz alta de listas de palabras, de frases, o de un párrafo escrito en prosa, al habla espontánea o libre, o a la combinación de ambos métodos); las escalas utilizadas (más comúnmente, escalas de *Likert* de diferentes niveles, desde los 3 o los 4 a los 9, o simplemente categorizaciones de nativo/no-nativo); los nativos de control (desde la proporción entre los no nativos y los nativos de control a la

información aportada sobre sus características y su historia lingüística); y, por último, los jueces o evaluadores (con una gran variabilidad, como veremos a continuación, en su número y características, y en la información detallada sobre ellos). (Ver Jesney, 2004, y Schmid y Hopp, 2014, para una revisión exhaustiva sobre todas estas cuestiones).

Obviamente, y en especial observando la enorme variabilidad existente,

> Any researcher wishing to conduct a foreign accent rating experiment will have to give consideration to a number of methodological issues concerning the material to be rated, the speakers to be rated, the raters and the rating scales. In many previous studies, these decisions appear to have been informed largely by convenience and are often not explicitly motivated or even reported, even though there is increasing evidence to show that seemingly minor variations in the setup can impact the outcome. (Schmid y Hopp, 2014, p. 369)

En este sentido, explicitar y reflexionar sobre los procedimientos seguidos, para tratar de establecer algún tipo de estandarización en torno a estas cuestiones metodológicas sobre informantes, materiales, escalas de evaluación y evaluadores utilizados es, como mínimo, deseable, y debería ser, de hecho, la praxis habitual en pos del rigor científico y de la verdadera posibilidad de falsación, réplica y comparación entre los resultados, a la que todos los agentes involucrados deberían -deberíamos- aspirar.

La selección de los jueces utilizados para la evaluación del grado de acento extranjero es, precisamente en esta línea, una de las cuestiones que, dada la variabilidad existente en la literatura, se antoja imprescindible estandarizar, delimitando las características más adecuadas que deberían reunir para establecer su validez y fiabilidad a la hora de diseñar la propia investigación.

3. LA UTILIZACIÓN DE PANELES DE JUECES PARA LA EVALUACIÓN DEL GRADO DE ACENTO EXTRANJERO (GLOBAL): ¿QUÉ TIPOS DE JUECES ELEGIR?

La forma más común y extendida de evaluación del grado de acento extranjero es la utilización de paneles de jueces. Sin embargo, una vez más, los estudios difieren ampliamente en cuanto al número y las características de los jueces seleccionados.

En cuanto al número de jueces utilizados, los estudios varían desde cantidades muy reducidas (un único juez en Snow y Hoefnagel-Höhle, 1977; dos en el estudio de Oyama, 1976; o tres en el de Birdsong, 2003) a números mucho más altos (llegando a los 224 utilizados en el estudio de Anderson-Hsieh y Köhler, 1988). También se han utilizado números intermedios. Granena y Long (2013), por ejemplo, incluyeron en su estudio un panel de 12 jueces nativos para juzgar el grado de acento extranjero de los informantes, frente a los 8 de Nagle (2013). Schmid y Hopp (2014), por su parte, utilizaron un total de 130.

Más allá de los números, la selección individual de los jueces para las pruebas de evaluación es, como señala Jesney, "also an important consideration, though relatively little research has directly considered this issue" (2004, p. 7). En la amplia mayoría de los casos, se prefiere la selección de hablantes nativos, aunque algunos estudios han utilizado también no nativos, y comparado sus evaluaciones (Riney, Takagi y Inutsuka, 2005; Gallardo del Puerto, García Lecumberri y Gómez Lacabex, 2015; Major, 2007; 2010). En general, en los últimos años parece existir un consenso sobre la conveniencia de limitarse a jueces nativos (ver Jesney, 2004).

Otro aspecto señalado en la bibliografía es el de la familiaridad de los jueces con las variedades, dialectos o acentos regionales que pueden aparecer en las muestras de habla evaluadas (Flege, Frieda y Nozawa, 1997; Bongaerts, van Summeren, Planken y Schils, 1997). Los resultados apuntan a que es conveniente tener, al menos, este aspecto en cuenta, ya que la disparidad de dialectos o variedades entre los informantes y los jueces puede suponer un inconveniente para el reconocimiento de la pronunciación nativa.

Por último, una cuestión aún controvertida es la de la utilización de jueces nativos expertos vs. inexpertos o *naïves*. ¿Deben ser jueces expertos, familiarizados con hablantes extranjeros, con entrenamiento o formación específica en lingüística y/o en fonología? ¿O deberíamos, más bien, confiar en jueces lingüísticamete inexpertos (n*aïves),* sin conocimientos específicos de lingüística, fonética, sin experiencia de familiaridad con hablantes de las L1 de los informantes del estudio?

Algunos autores proponen que, dada su formación específica, lingüistas y profesores de LE "are better FA perceivers than ordinary individuals, precisely because of their special training" (Gallardo del Puerto, García Lecumberri y Gómez Lacabex, 2015, p. 205). Así, los resultados del estudio de Thompson (1991) arrojaban que "experienced raters were more reliable in estimating the degree of accentedness across samples than were the less experienced raters" (p.199). Así también Calloway (1980) y Flege (1984) encontraron mayor fiabilidad en los jueces lingüísticamente expertos a la hora de evaluar el grado de acento extranjero. Por otro lado, se ha sugerido que los jueces expertos pueden presentar un mayor nivel de tolerancia hacia las desviaciones respecto a la norma nativa, resultando menos estrictos en sus valoraciones:

"language experts make sympathetic listeners" (Thompson 1991, p. 198). Otros estudios, sin embargo, apuntan a lo contrario, defendiendo que los jueces expertos constituyen evaluadores más críticos que los inexpertos (Schairer, 1992). Existen, no obstante, estudios que, tras comparar las evaluaciones realizadas por jueces expertos y por jueces inexpertos, no encontraron diferencias estadísticamente significativas entre unas y otras, sosteniendo así la idea de que los jueces *naïves* pueden ser tan fiables a la hora de evaluar el grado de acento extranjero como los expertos (Flege y Fletcher, 1992; Bongaerts, van Summeren, Planken & Schils, 1997; Anderson- Hsieh y Koehler, 1988; Uzal, Peltonen, Huotilainen y Aaltonen, 2015; Cunningham-Andersson y Engstrand, 1989; Brennan y Brennan, 1981)

Ante la diversidad de conclusiones presente en la bibliografía, y a falta de noticias de la existencia de estudios que hayan analizado esta cuestión con el par de lenguas francés L1/español L2,

el presente estudio se propuso comparar las evaluaciones del grado de acento extranjero en producciones orales (grabación de lectura de un párrafo en voz alta) en aprendices francófonos de español L2 realizadas por jueces expertos y por jueces inexpertos, tratando de arrojar luz sobre la fiabilidad de uno y otro tipo de jueces a la hora de llevar a cabo este tipo de investigaciones.

4. ESTUDIO EMPÍRICO

4.1 Breve descripción de la metodología:

En el presente estudio, centrado en la selección de jueces nativos para la evaluación del acento extranjero en aprendices de español L2, se plantearon las siguientes preguntas de investigación: (1)¿Resultan fiables los jueces inexpertos (o *naïves)* y los jueces expertos a la hora de reconocer el habla nativa? (2) ¿Cuál es el grado de fiabilidad interna de los jueces inexpertos (o *naïves*) y expertos en sus evaluaciones del grado de acento extranjero? (3) ¿Existe una diferencia significativa en la evaluación del acento extranjero por jueces inexpertos (o *naïves*) y por jueces expertos?

Los informantes seleccionados para el estudio fueron dos grupos de jueces elegidos a través de un muestreo no probabilístico por cuotas. El grupo de jueces nativos expertos se compuso de 12 profesores de ELE (7 mujeres y 5 hombres) con variedad andaluza y del centro-norte de España (de acuerdo con las dos variedades del español que podían aparecer en las muestras de habla evaluadas). El grupo de jueces nativos inexpertos lo formaban 12 nativos (5 hombres y 7 mujeres) sin ninguna formación específica en lingüística ni familiaridad con la lengua francesa, hablantes de las dos mismas variedades del español. Las grabaciones evaluadas por ambos paneles de jueces eran de 20 francófonos (8 hombres y 12 mujeres) residentes en España (un mínimo de 7 años) y con distintas edades de inicio de adquisición del español. Se añadió un grupo de control de 6 nativos (3 mujeres y 3 hombres) de variedad andaluza y del centro-norte de España en la muestra, para poder establecer la fiabilidad de los jueces.

La muestra de las grabaciones se componía, por tanto, de un total de 26. Para recoger estas muestras de habla, se diseñó un texto a propósito (modificado a partir del texto utilizado por Granena y Long, 2013) y se le pidió a los participantes que lo leyeran en voz alta para su grabación. Los participantes también completaron un cuestionario de histórico lingüístico sobre su experiencia de inmersión en España.

En cuanto a la evaluación de las grabaciones por los 24 jueces, se les informó de que en las grabaciones escucharían tanto a hablantes nativos del español como a no nativos, pero en ningún caso se les proporcionó ninguna información adicional sobre el porcentaje de nativos y no- nativos. En las instrucciones se les indicó que evaluasen el grado de acento extranjero de la pronunciación de los informantes en una escala del 0 al 10 (escala de puntuación habitual en el sistema de evaluaciones español), siendo 0 "acento extranjero muy fuerte" y 10 "sin acento extranjero" (= nativo hispanohablante). Cada juez evaluó cada una de las grabaciones (26 en total: 6 nativos de control y 20 aprendices francófonos) dos veces, en dos bloques separados, en cada uno de los cuales las grabaciones fueron distribuidas aleatoriamente. En cada bloque los jueces escucharon una única vez las grabaciones, con un tiempo de 5 segundos entre una y otra para poder anotar su puntuación. Antes de comenzar la prueba de evaluación, se presentó a los jueces una muestra de tres grabaciones no incluidas en la muestra para que les sirviesen como ejemplo del funcionamiento de la prueba.

4.2 Resultados

Para contestar a la primera pregunta de investigación planteada (¿Resultan fiables los jueces inexpertos y los jueces expertos a la hora de reconocer el habla nativa?) se decidió excluir a los jueces del panel cuando la desviación de la puntuación media de un juez a un informante nativo respecto de la media de las puntuaciones de todos los jueces fue mayor que la desviación típica. Se consideró entonces que el juez no resultaba fiable a la hora de reconocer una pronunciación nativa del español. Como puede apreciarse en las

tablas 1 y 2 (a continuación), 4 jueces expertos y 3 jueces inexpertos fueron eliminados del panel final de jueces siguiendo este criterio.

Jueces Expertos	Nativo 1	Nativo 2	Nativo 3	Nativo 4	Nativo 5	Nativo 6	Decisión
Juez P1	10	10	10	10	10	10	Se mantiene
Juez P2	10	10	10	10	10	10	Se mantiene
Juez P3	10	10	10	10	10	9	Se elimina
Juez P4	10	10	10	10	10	10	Se mantiene
Juez P5	10	10	10	10	10	10	Se mantiene
Juez P6	9.50	8.50	10	10	8	10	Se elimina
Juez P7	10	10	10	10	10	10	Se mantiene
Juez P8	9.50	10	10	10	10	10	Se mantiene
Juez P9	10	10	10	9.50	10	10	Se elimina
Juez P10	9.50	10	9	10	10	9.50	Se elimina
Juez P11	10	10	10	10	10	10	Se mantiene
Juez P12	10	10	10	10	10	10	Se mantiene
Media Global	9.77	9.94	9.81	9.90	9.85	9.85	
Desv. Típ	0.63	0.31	0.41	0.33	0.45	0.38	
Fiabilidad	**9.14**	**9.63**	**9.4**	**9.57**	**9.4**	**9.47**	

Tabla 1: puntuaciones nativos jueces expertos (criterio fiabilidad 1)

Jueces Naïve	Nativo 1	Nativo 2	Nativo 3	Nativo 4	Nativo 5	Nativo 6	Decisión
Juez N1	9.50	10	10	10	10	10	Se mantiene
Juez N2	10	10	10	10	10	10	Se mantiene
Juez N3	10	10	10	10	10	10	Se mantiene
Juez N4	10	10	10	10	10	10	Se mantiene
Juez N5	10	10	10	10	10	10	Se mantiene
Juez N6	10	10	9	10	10	10	Se elimina
Juez N7	10	10	9.50	10	10	10	Se mantiene
Juez N8	10	10	10	10	10	10	Se mantiene
Juez N9	9.50	10	9.50	9.50	9.50	10	Se elimina
Juez N10	10	10	10	10	10	9.50	Se mantiene
Juez N11	10	10	10	10	10	10	Se mantiene
Juez N12	7	10	8.50	8.50	9	8.50	Se elimina
Media Global	9.81	9.94	9.83	9.92	9.87	9.87	
Desv. Típ.	0.58	0.40	0.47	0.32	0.48	0.40	
Fiabilidad	**9.14**	**9.63**	**9.4**	**9.57**	**9.4**	**9.47**	

Tabla 2: puntuaciones nativos jueces naives (criterio fiabilidad 1)

Para el segundo criterio de fiabilidad, correspondiente a la segunda pregunta de investigación (¿Cuál es el grado de fiabilidad interna de los jueces inexpertos y expertos en sus evaluaciones del grado de acento extranjero?) se calculó el coeficiente de fiabilidad intra-jueces. Para ello, los jueces evaluaron 2 veces a cada informante y para cada uno se calculó el coeficiente de correlación entre las dos series de puntuaciones (evaluación 1 y evaluación 2). Se mantuvo únicamente a aquellos jueces que obtuvieron una correlación positiva significativa alta o muy alta. El resto fue descartado de los paneles

de jueces finales. Siguiendo este criterio, únicamente un juez fue eliminado, perteneciente al grupo de expertos, por presentar un coeficiente de correlación demasiado bajo (0.583 –correlación positiva significativa *moderada*-) de acuerdo con el criterio utilizado (correlación positiva significativa alta o muy alta). (Tablas 3 y 4, a continuación)

Jueces Expertos	Correlación entre las 2 evaluaciones	Interpretación	Decisión
P1	0.947	Muy alta	Se mantiene
P2	0.938	Muy alta	Se mantiene
P4	0.874	Alta	Se mantiene
P5	0.964	Muy alta	Se mantiene
P7	0.852	Alta	Se mantiene
P8	0.800	Alta	Se mantiene
P11	0.838	Alta	Se mantiene
P12	**0.583**	**Moderada**	Se elimina

Tabla 3: Coeficiente de fiabilidad intra-jueces (JUECES EXPERTOS)

Jueces *Naïves*	Correlación entre las 2 evaluaciones	Interpretación	Decisión
N1	0.775	Alta	Se mantiene
N2	0.835	Alta	Se mantiene
N3	0.891	Alta	Se mantiene
N4	0.978	Muy alta	Se mantiene
N5	0.808	Alta	Se mantiene
N7	0.839	Alta	Se mantiene
N8	0.970	Muy alta	Se mantiene
N10	0.838	Alta	Se mantiene
N11	0.932	Muy alta	Se mantiene

Tabla 4: Coeficiente de fiabilidad intra-jueces (JUECES NAÏVES)

Para contestar a nuestra tercera pregunta de investigación (¿Existe diferencia significativa entre el grupo de jueces expertos y el grupo de jueces inexpertos?), se plantearon las siguientes hipótesis: (Ho) No existe diferencia significativa entre el grupo de jueces expertos y el grupo de jueces inexpertos; (Ha) Existe diferencia significativa entre el grupo de jueces expertos y el grupo de jueces inexpertos. Después de haber comprobado la

normalidad en la distribución de los datos, se aplicó la prueba paramétrica *T-Test* de muestras independientes. Como puede apreciarse en la tabla 5, no se pudo rechazar la hipótesis nula, por lo que podemos concluir que la diferencia entre ambos grupos no es significativa (p=0.353 (>0.05):

Prueba de muestras independientes

		Prueba de Levene de igualdad de varianzas		prueba t para la igualdad de medias						
									95% de intervalo de confianza de la diferencia	
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Diferencia de error estándar	Inferior	Superior
MediaPN	Se asumen varianzas iguales	1,976	,166	,938	50	,353	,45220	,48208	-,51608	1,42048
	No se asumen varianzas iguales			,938	47,920	,353	,45220	,48208	-,51713	1,42152

Tabla 5: Resultados prueba *T-Test*

5. CONCLUSIONES

El análisis de estos resultados nos permite concluir, en primer lugar, que no existió diferencia estadísticamente significativa entre la evaluación de los jueces inexpertos y los expertos. Nuestros resultados coinciden, en este sentido, con los de estudios previos, que señalaron notables similitudes entre las evaluaciones de estos tipos de jueces (Flege y Fletcher, 1992; Bongaerts, van Summeren, Planken & Schils, 1997; Anderson-Hsieh y Koehler, 1988; Uzal et al., 2015; Cunningham-Andersson y Engstrand, 1989; Brennan y Brennan, 1981). Los jueces inexpertos o *naïves* mostraron una capacidad similar a la de los expertos a la hora de evaluar el acento extranjero, a pesar de no tener ninguna formación en lingüística y de no tener una relación de familiaridad con hablantes de otras lenguas. Cabe destacar, en este sentido, que, atendiendo a ambos criterios de fiabilidad, los jueces inexpertos resultaron ligeramente más fiables que los expertos. Mediante el análisis correlacional, se comprueba, de hecho, que los jueces inexpertos resultaron incluso más fiables (fiabilidad intra-jueces) que los expertos. Asimismo, el 75% de ellos fue capaz de reconocer a los nativos en el panel de muestras, mientras solo un 67 % de los jueces expertos lo consiguieron.

Aunque no exista aún un consenso sobre el tipo de jueces que se debe utilizar para la evaluación del acento extranjero en este tipo de estudios, queremos subrayar que este tipo de análisis de

fiabilidad y de procedimientos se revelan imprescindibles para evitar la inconsistencia en las puntuaciones de los evaluadores. Asimismo, consideramos que compartir prácticas para los procedimientos de selección de jueces contribuirá a avanzar en esta área de investigación. Los resultados de este estudio aportan, en esta línea, argumentos para apoyar la selección de jueces inexpertos, que se basarán, probablemente, para sus evaluaciones, en su intuición lingüística más que en sus conocimientos teóricos sobre el tema. No obstante, sería necesario, para confirmar estos resultados, ampliar el tamaño de las muestras, tanto de las grabaciones como de los dos paneles de jueces, para disponer así de resultados más consecuentes. Podrían también, en esta misma línea, estudiarse los coeficientes de fiabilidad inter-jueces y profundizar así también en el análisis de la fiabilidad de la herramienta de evaluación. En ese sentido, resultaría igualmente de interés llevar a cabo un estudio contrastivo con distintas escalas de evaluación, para reflexionar en profundidad sobre el número de puntuaciones más acertado para dichas escalas.

REFERENCIAS BIBLIOGRÁFICAS

Abrahamsson, 2012. "Age of onset and nativelike L2 ultimate atainment in morphosyntactic and phonetic intuition". *Studies in Second Language Acquisition*, 34: 187-214.

Abrahamsson, N. y Hyltenstam, K., 2009. "Age of Onset and Nativelikeness in a Second Language: Perception versus Linguistic Scrutiny". *Language Learning*, 59 (2): 249-306.

Anderson-Hsieh, J. y Köhler, K., 1988. "The effect of foreign accent and speaking rate on native speaker comprehension". *Language Learning*, 38: 561-593.

Asher, J. T. y García, R., 1969. "The Optimal Age to Learn a Foreign Language". *The Modern Language Journal*, 53 (5): 334-341.

Birdsong, D., 2007. "Nativelike pronunciation among late learners of French as a second language". En: O.-S. Bohn y M. Munro, eds. *Language Experience in Second Language Speech Learning*. Amsterdam: John Benjamins: 99-116.

Birdsong, D., 2003. "Authenticité de prononciation en français L2 chez des apprenants tardifs anglophones: Analyses segmentales et globales". *Acquisition et Interaction en Langue Étrangère*, 18: 17-36.

Bongaerts, T., 1999, "Ultimate attainment in L2 pronunciation: The case of very late advanced late L2 learners". En: *Second Language Acquisition and the Critical Period Hypothesis*. Mahwah, NJ: Erlbaum: 133-159.

Bongaerts, T., Mennen, S. y van der Slik, F., 2000. "Authenticity of pronunciation in naturalistic second language acquisition: the case of very advanced late learners of dutch as a second language". *Studia Linguistica*, 54 (2): 298-308.

Bongaerts, T., van Summeren, C., Planken, B. y Schils, E., 1997. "Age and ultimate attainment in the pronunciation of a foreign language". *Studies in Second Language Acquisition*, 19: 447-465.

Bongaerts, T., van Summeren, C., Planken, B. y Schils, E., 1997. "Age and ultimate attainment in the pronunciation of a foreign language". *Studies in Second Language Acquisition*, 19: 447-465.

Brennan, E. M. y Brennan, J. S., 1981. "Measurements of Accent and Attitude Toward Mexican- American Speech". *Journal of Psycholinguistic Research*, 10(5): 487-501.

Calloway, D. R. (1980) Accent and the evaluation of ESL oral proficiency. En: J. W. Oller Jr. and K. Perkins (eds.) 1980 *Research in language testing*. Rowley, MA: Newbury House: 102–15.

Colantoni, L. y Steele, J., 2006. "Native-like attainment in the L2 acquisition of Spanish stop- liquid clusters". En: C. A. Klee y T. L. Face, eds. *Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages*. Somerville, MA: Cascadilla Proceedings Project: 59-73.

Cunningham-Andersson, U. y Engstrand, O., 1989. "Perceived strength and identity of foreign accent in Swedish". *Phonetica*, 46: 138-154.

DeKeyser, R. M., 2013. "Age Effects in Second Language Learning: Stepping Stones Toward Better Understanding". *Language Learning*, 63 (1): 52-67.

Flege, J. E., 1999. "Age of Learning and Second-language Speech". En: Birdsong, D., ed.

1999. *Second Language Acquisition and the Critical Period Hypothesis*. Hillsdale, NJ: Lawrence Erlbaum: 101-132

Flege, J. E., 1988. "Factors affecting degree of perceived foreign accent in English sentences".

Journal of Acoustical Society of America, 84: 70-79.

Flege, J. E., 1984. "The detection of French accent by American listeners". *Journal of the Acoustical Society of America*, 76.3: 692–707.

Flege, J. E. y Fletcher, K. L., 1992. "Talker and listener effects on degree of perceived foreign accent". *Journal of the Acoustical Society of America*, 9(1): 370-389.

Flege, J. E., Frieda, E. M. y Nozawa, T., 1997. "Amount of Native-Language (L1) Use Affects the Pronunciation of an L2". *Journal of Phonetics*, 25: 169-186.

Flege, J. E., Munro, M. J. y MacKay, I. R. A., 1995. "Factors affecting strength of perceived foreign accent in a second language". *Journal of Acoustical Society of America*, 87 (5): 3125-3134.

Flege, J. E., Yeni-Komshian, G. H. y Liu, S., 1999. "Age Constraints on Second Language Acquisition". *Journal of Memory and Language*, 41: 78-104.

Gallardo del Puerto, F., García Lecumberri, M. L. and Gómez Lacabex, E., 2015. "The assessment of foreign accent and its communicative effects by naïve native judges vs. experienced non-native judges". *International Journal of Applied Linguistics*, 25: 202– 224.

Grandfeldt-Saavedra, P., 2007. "Adquisición de vocabulario y pronunciación en L1 y L2.

Estudio comparativo de los factores edad e *input* en el castellano de suecos bilingües". En: J. Grandfeldt, ed. *Studies in Romance Bilingual Acquisition: Age of Onset and Development of French and Spanish*. Lund: Lund University: 43-89.

Granena, G. y Long, M., 2013. "Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains". *Second Language Research*, 29 (3): 311-343.

Hyltenstam, K. y Abrahamsson, N., 2003. "Age de l' exposition initiale et niveau terminal chez des locuteurs quasi-natifs du suédois L2". *Acquisition et interaction en langue étrangère*, [en línea] 18. Disponible en línea a través de: revues.org website <http://aile.revues.org/1154> [Consultado 16 de enero de 2014].

Jesney, K. (2004). *The use of global foreign accent rating in studies of L2 acquisition*. Calgary, AB: Language Research Centre, University of Calgary.

Lenneberg, E., 1967. *Biological Foundations of Language*. New York: Wiley.

Long, M., 2005. "Problems with supposed counter-evidence to the Critical Period Hypothesis".

IRAL, 43: 287-317.

Major, R. C., 2007. "Identifying a foreign accent in an unfamiliar language". *Studies in Second Language Acquisition*, *29*: 539–556.

Major, R. C., 2010. "First language attrition in foreign language perception". *International Journal of Bilingualism*, *14*: 163–183.

Marinova-Todd, S. H., Marshall, D. B. y Snow, C. E., 2000. "Three Misconceptions about Age and L2 Learning". *TESOL Quarterly*, 34 (1): 9-34.

Moyer, A., 1999. "Ultimate Attainment in L2 Phonology. The Critical Factors of Age, Motivation and Instruction". *Studies in Second Language Acquisition*, 21: 81-108.

Nagle, C., 2013. "A Reexamination of Ultimate Attainment in L2 Phonology: Length of Immersion, Motivation, and Phonological Short-Term Memory". En: ed. Erik Voss et al., *Selected Proceedings of the 2011 Second Language Research Forum*, Somerville, MA: Cascadilla Proceedings Project: 148-161.

Oyama, S., 1976. "A sensitive period for the acquisition of a nonnative phonological system".
Journal of Psycholinguistic Research, 5 (3): 261-283.
Patkowski, M., 1990. "Age and accent in a second language: A reply to James Emil Flege".
Applied Linguistics, 11: 73-89.
Piske, T., MacKay, I. R. A. y Flege, J. E., 2001. "Factors affecting degree of foreign accent in an L2: a review". *Journal of Phonetics*, 29: 191-215.
Purcell, E. y Suter, R., 1980. "Predictors of pronunciation accuracy: A reexamination".
Language Learning, 30 (2): 271-287.
Riney, T. J., Takagi, N. y Inutsuka, K., 2005. "Phonetic Parameters and Perceptual Judgments of Accent in English by American and Japanese Listeners". *TESOL Quarterly*, 39 (3): 441- 466.
Schmid, M. S. y Hopp, H., 2014. "Comparing foreign accent in L1 attrition and L2 acquisition: Range and rater effects". *Language Testing*, 31 (3): 367–388.
Seliger, H., Krashen, S. y Ladefoged, P., 1975. "Maturational constraints in the acquisition of second language accent". *Language Science*s, 36: 20-22.
Schairer, K. E., 1992. "Native speaker reaction to non-native speech". *Modern Language Journal*
76.3: 309–19.
Snow, C. y Hoefnagel-Höhle, M., 1977. "Age differences in the pronunciation of foreign sounds". *Language and Speech*, 20: 357-365.
Suter, R., 1976. "Predictors of pronunciation accuracy in second language learning". *Language Learning*, 26 (2): 233-253.
Tahta, S., Wood, M. y Loewenthal, K., 1981a. "Foreign accents: Factors Relating to Transfer of Accent from the First Language to a Second Language". *Language and Speech*, 24 (3): 265-272.
Tahta, S., Wood, M. y Loewenthal, K., 1981b. "Age Changes in the Ability to Replicate Foreign Pronunciation and Intonation". *Language and Speech*, 24 (4): 363-373.

Thompson, I., 1991. "Foreign accents revisited: The English pronunciation of Russian immigrants". *Language Learning*, 41 (2): 177-204.

Uzal, M., Peltonen, T., Huotilainen, M. and Aaltonen, O., 2015. "Degree of Perceived Accent in Finnish as a Second Language for Turkish Children Born in Finland". *Language Learning*, 65: 477–503.